
THE UNIVERSITY
*of* EDINBURGH

# Investigating the Functional Role of SMOC-1 in Zebrafish

David Sexton

**Presented for the Degree of**

**Doctor of Philosophy**

The University of Edinburgh

October 2016

I declare that this thesis is composed of original research and analysis undertaken by myself, and where the work of others is included, their contributions have been duly acknowledged.

David Sexton

October 2014

# Acknowledgements

The first person thank is David Fitzpatrick. I couldn't have asked for a better and more understanding supervisor, particularly in the final year. I really appreciate all the time you put into training me and thank you for inviting me into the Fitzpatrick Lab. I'd also like to thank the other members of the Fitzpatrick lab. In particularly; Joe Rainger for identifying SMOC-1as a cause of OAS and providing lots of advice and reagents; Morad Ansari for your advice on ion torrent sequencing; Kishan Aldridge for your contribution to the early zebrafish work; Hemant Bengani for you advice with the protein work and Anne Seawright for your help on with developing the Southern blotting and your skills at managing the. I also have to thank both the van Heyningen and Jackson labs as they were excellent lab colleagues, both socially and technically.

The next group of people I'd like to thank is the IGMM fish community; most notably Elizabeth Patton who provided advice and motivation. It was a pleasure to work with your lab. My particular thanks goes out to Witold Rybski who is a star. Thanks for looking after my fish* and especially performing so many injections for this project, including all of the mRNA injections report in this thesis**. Many thanks to Corina, Kerry, Juan and Amy in the Patton lab who all provided me help, advice and and the ocassional reagent. Pedro and Sofia for being part of the Yamuna room, and Keith for your excellent care of the fish room.

Like many Ph.D. students in the Institute I have to offer my heartfelt thanks Paul Perry and Matt Pearson for their excellent services in the imaging unit and help in establishing collide zebrafish imaging. Particularly Matt who even returned my calls at the most unsociable of ours. Thanks to Harris Morrison I was able to do OPT imaging. He conducted the OPT scans and taught me how to use the relevant software***.

I'm grateful to the MRC for providing me and my research with funding****.

I'd also like to thank my family and friends. Particularly those who helped proof read my thesis. Thanks to everyone who supported me. An extra special thanks goes out to my partner Rachel, who has provided me with so much mental and practical support through writing up my Ph.D.

* General fish husbandry was performed by fish room technicians. Though it should be noted that line maintenance and genotyping was performed by myself

** All injections were conducted by Witold Rybski in Figures 3.2.3, 3.2.5, 3.2.6, 3.2.7, 3.2.9, 3.2.10, 3.2.14, 3.2.15. In all other instances the injects were performed by myself

*** All OPT scans were performed by Harris Morrison as detailed in **6.3.1.4**

**** Both ION torrent and Sanger sequencing was performed by (IGMM) sequencing service as detailed in sections 6.4.7 and 6.4.6 though analysis was conducted by myself.

# Abstract

True anophthalmia is the most severe congenital eye malformation. With absence of the eye, optic nerve, chiasm and optic tracts. Identifying the genes that cause genetic true anophthalmia should improve our understanding of the critical processes required for development of the eye. Recessive loss-of-function mutations in *SMOC1* have been identified as the cause of Ophthalmo-acromelic syndrome (OAS), a multisystem disorder which has true anophthalmia as a prominent feature with characteristic limb and facial malformations. In order to establish the function of *SMOC1* in development I used the zebrafish as a model organism to support a link between SMOC-1 and BMP signalling.

As a first step I characterised the genomic structure of zebrafish *smoc1* gene. I was able to correct an error in the zebrafish genome (Zv8) that annotated *zsmoc1* as two fragmented and rearranged orthologous loci. However, using RT-PCR I could show that there is in fact a single intact z*smoc1* transcript. In addition, I was able to identify an un-annotated 5' coding exon using 5' RACE which showed that the full open reading frame includes a signalling peptide. RT-PCR was also used to identify several novel *zsmoc1* splice isoforms.

To explore the link between *zsmoc1* and bmp signalling I used injection of antisense morpholino oligonucleotide and capped mRNA to examine the effects of loss-of-function and overexpression respectively of *smoc1* and genes functioning in

the bmp signalling pathway. The resulting embryos were analysed using morphometric analysis (Kishimoto scale), a quantitative assay of dorsalisation/ventralisation and live imaging of reporter transgenic fish. I developed a quantitative RT-PCR assay for expression of dorsal (*otx2* and *runx3)* and ventral (*eve1* and *gata2*) marker genes. I established a reliable system for live imaging of zebrafish development between 8 hpf and 24 hpf. By combining this system with fluorescent transgenic reporters marking the eye field (*rx3:gfp* reporter) and BMP-signaling (BRE:gfp reporter) I was able to accurately quantitate the effect of smoc1 depletion on eye size and SMAD1/5/8 signalling in the eye. These results support the predictions from the *Drosophila* homologue *Pent* that *zsmoc1* functions as an antagonist of bmp signalling.

Finally, I describe my attempt to produce a zebrafish model for OAS using genome editing technology. I designed, produced and validated transcription activator like effectors nucleases (TALENs) targeted to the *zsmoc1* open reading frame using the Voytas Goldengate method. I designed and optimised a novel strategy to demonstrate targeted cutting activity for *in vitro* validation. Following injections of the *in vitro* validated TALEN into zebrafish embryos I used Ion Torrent sequencing to assess the *in vivo* activity of the engineered TALEN pairs. Unfortunately these TALENs were not able to cut the targeted locus *in vivo*.

# Lay Summery

True anophthalmia is the most severe congenital eye malformation with absence of the eye, optic nerve, chiasm and optic tracts. Identifying the genes that cause genetic true anophthalmia should improve our understanding of the critical processes required for development of the eye. Recessive loss-of-function mutations in *SMOC1* have been identified as the cause of Ophthalmo-acromelic syndrome (OAS), a multisystem disorder which has true anophthalmia as a prominent feature with characteristic limb and facial malformations. In order to establish the function of *SMOC1* in development I used the zebrafish as a model organism to support a link between SMOC1 and BMP signalling.

As I first step characterised the genomic structure of zebrafish *smoc1* gene. I was able to correct an error in the zebrafish genome (Zv8) that annotated two fragmented and rearranged orthologous loci. However, using RT-PCR I could show that there is in fact a single intact *smoc1* transcription unit. In addition I was able to identify an un-annotated 5' coding exon using 5' RACE which showed that the full open reading frame includes a signalling peptide. RT-PCR was also used to identify several novel *zsmoc1* splice isoform.

To explore the link between *zsmoc1* and BMP signalling I used injection of antisense morpholino oligonucleotide and capped mRNA to examine the effects of loss-of-function and over-expression respectively of *smoc1* and genes functioning

in the BMP signalling pathway. The resulting embryos were analysed using morphometric analysis (Kishimoto scale), a quantitative assay of dorsalisation/ventralisation and live imaging of reporter transgenic fish. I developed the qualtitative RT-PCR assay for expression of dorsal (*otx2* and *runx)* and ventral (*Eve1* and *gata2*) marker genes. I established a reliable system for live imaging of zebrafish development between 8 hpf and 36 hpf. By combining this system with fluorescent transgenic reporters marking the eye field (*rx3*) and BMP-signaling (BRE reporter) I was able to accurately quantitate the effect of smoc1 depletion on eye size and SMAD1/5/8 signalling in the eye. These results support the predictions from the drosophila orthology that zsmoc1 functions as an antagonist of BMP signalling.

Finally I describe my attempt to produce a zebrafish model for OAS using genome editing technology. I designed, produced and validated transcription activator like effectors nucleases(TALENs) targeted to the *zsmoc1* open reading frame using the Voytas Goldengate method. I designed and optimised a novel strategy to demonstrate targeted cutting activity *in vitro* validation. Following injections of the in vitro validated TALEN into zebrafish embryos I used ion torrent sequencing to assess the in vivo activity of the engineered TALEN pairs. Unfortunately these TALENs were not able to cut the targeted locus *in vivo*.

# Table of contents

**LIST OF FIGURES ..... 17**

**LIST OF TABLES ..... 22**

**ABBREVIATIONS ..... 23**

**INTRODUCTION ..... 30**

1.1 Ophthalmo-acromelic syndrome ..... 31

1.1.1 Clinical features ..... 31

1.1.2 The identification of mutations in *SMOC1*, the principal genetic cause of OAS ..... 33

1.1.3 The structure of SMOC-1 ..... 34

1.1.4 The mode of action of Smoc-1 in animal models ..... 36

1.2 Eye development ..... 41

1.2.1 Vertebrate eye development ..... 41

1.2.2 Known genetic causes of anophthalmia and microphthalmia ..... 49

1.2.2 Potential causes of anophthalmia ..... 58

1.3. Main Aims ..... 60

Chapter 4 ..... 65

**CHARACTERISING THE *SMOC1* LOCUS IN ZEBRAFISH ..... 66**

2.1 Introduction ..... 67

2.1.1 *Danio rerio* ..... 67

2.1.2 *Danio rerio* as a useful model of disease ..... 68

2.1.3 Zebrafish Genomics ..... 71

2.2 Results ..... 75

2.2.1 Clarifying the Genomic organization of *zsmoc1* ..... 75

2.2.2 The recorded sequence of zebrafish *zsmoc-1* is incomplete ..... 77

2.2.3 5'RACE identifies the signal peptide open reading frame and 5' UTR of *zsmoc1* ..... 79

2.2.4 Identifying transcript variants of *zsmoc1* ..... 82

2.2.5 The temporal expression pattern of *zsmoc1* transcript variants ..... 88

2.2.6 Identification of candidate novel splice sites in human *SMOC1* ..... 89

2.2.7 Functional predictions using *zsmoc1* isoforms ..... 93

2.3 Discussion ..... 97

2.3.1 The zebrafish genomic locus has been established ..... 97

2.3.2 Alternative splicing in *zsmoc1* action ..... 98

2.3.3 Cross species analysis as a means of detecting alternative transcripts in humans ..... 101

2.3.4 Concluding remarks ..... 104

**ZSMOC-1: AN ANTAGONIST OF BMP SIGNALLING.....105**

3.1 Introduction.....106

3.1.1 Developmental signalling.....106

3.1.2 The Bmp signalling cascade.....107

3.1.3 Bmp signalling gradients.....112

3.1.4 bmp gradients in zebrafish development: dorsoventral patterning of the zebrafish embryo and eye.....118

3.1.5 smoc-1 is linked BMP signalling.....125

3.1.6 The current evidence for a link between smoc1 and bmp signalling in zebrafish.....127

3.2 Results.....130

3.2.1 Characterising zebrafish *smoc1* (*zsmoc1*) morphant phenotype.....130

3.2.2 Developing a quantitative RTPCR assay to assess dorsalisation and ventralisation in morphant zebrafish embryos.....132

3.2.3 The effect of zsmoc-1 depletion on ventral ectoderm markers and dorsal neural markers.....136

3.2.4 The effect of *zsmoc1* mRNA injection on ventral ectoderm markers and dorsal neural markers.....138

3.2.5 *hSMOC1* mRNA injection partially rescues zsmoc1MO injection.....141

3.2.6 Morpholino knockdown of *bmpr2b* gives similar results to morpholino knockdown of *zsmoc1*.....143

3.2.7 *hSMOC1* mRNA does not rescue either *bmp2b* mRNA induced ventralisation or bmp2b MO induced dorsalisation phenotypes ..... 144

3.2.7 Optimising automated live imaging protocol for early zebrafish embryos ..... 147

3.2.8 The effect of *zsmoc1* depletion on the developing zebrafish eye ..... 150

BMP signalling in the developing zebrafish eye ..... 154

3.2.10 *zsmoc1* morphants exhibit and a laterally expanded dorsoventral p-smad gradient at 60% epiboly(6hpf) ..... 157

3.3 Discussion ..... 160

3.3.1 zsmoc-1, an antagonist of BMP signalling? ..... 160

3.3.2 A model for the feedback between zsmoc-1 and bmp signalling ..... 164

3.3.3 *zsmoc1* depletion does not result in a morphological change in eye development ..... 173

3.3.4 Concluding remarks ..... 180

**ESTABLISHING AND STREAMLINING THE PRODUCTION OF TALE NUCLEASES ..... 182**

4.1 Introduction ..... 183

4.1.1 Traditional genomic editing in zebrafish ..... 183

4.1.2 TALEN proteins ..... 185

4.1.3 TALEN synthesis strategies ..... 188

2.1.4 Aims ..... 193

4.2 Results ..... 194

4.2.1 TALEN design and targeting ..... 194

4.2.2 TALEN construction ..... 199

4.2.3 In vitro confirmation of TALEN endonuclease activity ..... 209

4.2.4 Assessing the endonucleic activity of TALEN pairs 1 and 2 on *zsmoc1* in vivo ..... 215

4.3 Discussion ..... 228

4.3.1 Neither TALEN pair produced cleavage in vitro. ..... 228

4.3.2 Evaluating the novel assays developed to assess TALEN activity ..... 231

4.3.3 Alternatives to a TALEN based strategy ..... 233

4.3.4 Concluding remarks ..... 237

**DISCUSSION ..... 280**

5.1 Future work ..... 239

5.1.1 Generating a *zsmoc1*$^{-/-}$ zebrafish lines ..... 239

5.1.2 Investigate whether zsmoc-1 is acting on any other signalling pathways ..... 241

5.1.3 Establish the protien localisation of zsmoc1 ..... 243

5.1.4 Investigate the link between zsmoc-1 action, bmp signalling and the glypicans ..... 243

5.2. Contribution to the field ..... 246

5.2.1 Corrected the zebrafish locus of the zsmoc1 ..... 246

5.2.2 Established novel assays of cleavage efficiencies of targeted in endonucleases ..... 247

5.2.3 Provided the first functional work supporting a model where smoc-1 acts as a regulator of gradient shape ..... 247

**BIBLIOGRAPHY ..... 249**

**METHODS ..... 280**

6.1 Zebrafish husbandry ..... 284

6.1.1 Embryo care ..... 284

6.1.2 Embryo Bleaching ..... 285

6.1.3 Dechorionation ..... 287

6.1.4 Anaesthesia and Termination ..... 288

6.2 Microinjections ..... 289

6.2.1 Morpholino and mRNA solutions ..... 289

6.2.2 Needle preparation ..... 290

6.2.3 Microinjections ..... 292

6.2.4 Phenotyping ..... 292

6.3. Imaging ..... 293

6.3.1 Live zebrafish imaging ..... 293

6.3.2 Gradient quantification with OPT ..... 297

6.4 Nucleic acid techniques ..... 299
6.4.1 General methods and analysis ..... 299
6.4.2 DNA purification ..... 300
6.4.3 RNA purification ..... 303
6.4.4 Spectrophotometric analysis ..... 306
6.4.5 Gel electrophoresis ..... 306
6.4.6 Sequencing ..... 307
6.4.7 Ion Torrent sequencing ..... 308
6.4.8 Polymerase chain reaction PCR ..... 311
6.4.9 RACE PCR ..... 317
6.5 Protein techniques ..... 319
6.5.1 In vitro synthesis ..... 319
6.5.2 Western Blotting ..... 320
6.5.3 In vitro digestions of DNA using TALEN pairs ..... 325
6.5.4 Southern Blotting ..... 326
6.5.5 Whole mount immunohisto-chemistry ..... 327
6.5.6 Bioinformatic analysis ..... 329
6.6 Microbial Methods ..... 331
6.6.1 Basic microbial conditions ..... 331
6.6.2 Sub-cloning ..... 332

6.6.3 Transformation ..... 336

**SUPPLEMENTARY FIGURES ..... 337**

Supplementary Chapter 1 ..... 338

Supplementary Chapter 2 ..... 339

Supplementary Chapter 4 ..... 343

Supplementary methods ..... 345

Supplementary Nucleic acid techniques ..... 345

Supplementary Protein Techniques ..... 351

Supplementary Microbial Methods ..... 352

# List of Figures


| Figure Title | Page |
|---|---|
| **Fig. 1.1.1** Ophthalmo-acromelic syndrome and Smoc1 tm1a/tm1a mice are both characterised by bilateral eye and limb malformations. | 32 |
| **Fig 1.1.2** The a side by side comparison of the structure of the SMOC subfamily of BM40 protein and related *Drosophila* homologue Pentagone. | 34 |
| **Fig 1.1.3** The known expression patterns of SMOC-1 during deve-lopment. | 37 |
| **Fig 1.1.4** The Vuilleumier model of pentagone action in *Drosophila* | 39 |
| **Fig. 1.2.1** Schematic of the development of a generic vertebrate eye. | 42 |
| **Fig. 1.2.2** The sequence of development for the vertebrate eye. | 43 |
| **Fig. 1.2.3** Schematic of the expression of BMP signalling proteins in dorsoventral patterning of the vertebrate eye. | 46 |
| **Fig. 2.1.1** Schematic representation of different outcomes of genomic assembly. | 73 |
| **Fig. 2.2.1** The genomic organisation of the zebrafish *smoc1* gene. | 76 |
| **Fig. 2.2.2** A comparison of signal peptide prediction algorithms on SMOC-1 in multiple species. | 78 |
| **Fig. 2.2.3** 5'RACE identified a novel 5' UTR and exon coding the signal peptide of *zsmoc1*. | 81 |
| **Fig. 2.2.4** The structure of the zebrafish *smoc1* gene and its most common isoforms. | 84 |
| **Fig. 2.2.5.** The structure of the zebrafish *smoc1* gene and its minority isoforms. | 86 |

**Fig. 2.2.6.** The temporal expression patterns of the major *zsmoc1* variants in early development. 87
**Fig. 2.2.7.** An activity flow diagram detailing the search strategy for the cross species analysis of *zsmoc1* and *hSMOC1* alternative splicing. 90

**Fig 2.2.8.** Multiple sequence alignments of *Smoc1* across four species. 92

**Fig. 2.2.9** Splicing paterns are conserverd between human SMOC1 and zebrafish smoc1. 93

**Fig 2.2.9.** In silico analysis of the structure of *smoc1* for the alternative splice variants. 95

**Fig 3.1.1.** The canonical in BMP signalling pathway. 108

**Fig 3.1.2** zebrafish gastrulation. 119

**Fig 3.1.3.** The Kishimoto scale of zebrafish dorsalisation. 120

**Fig 3.1.4.** Antero-posterior and dorsoventral patterning in zebrafish embryos requires smoc1. 128

**Fig 3.2.1** New *zsmoc1* targeting morpholinos induce the characteristic dorsalisation phenotypes previously reported. Even when co-inject with p53 targeting MO. 131

**Fig 3.2.2.** Down-regulation of bmp signalling in *bmp2b* morphants leads to the predicted reduction in the expression of ventral ectoderm markers and a corresponding up-regulation of *smoc1* and neural markers at 14hpf but not at 9hpf. 133

**Fig. 3.2.3** Upregulation of bmp signalling leads to an up-regulation of ventral ectoderm markers and a corresponding down-regulation of *zsmoc1* and neural markers at 14hpf. 135

**Fig. 3.2.4.** The neural and ventral ectoderm marker genes are universally down regulated at 14 hpf *zsmoc1* morphants. 137

**Fig. 3.2.5** *zsmoc1* mRNA injection does not significantly alter embryo patterning. 139

**Fig. 3.2.6** *hSMOC1* mRNA injection acts leads to defects in embryo patterning consistent with *zsmoc1* acting like an agonist 140

of BMP signalling.

**Fig. 3.2.7** Injection of *hSMOC1* mRNA can partially rescue *zsmoc1* morphants. 142

**Fig. 3.2.8** *bmpr2a* and *bmp2a* morpholinos act synergistically to control dorsoventral patterning. 143

**Fig. 3.2.9** Co-injection *hSMOC1* mRNA does not alter the expression of neural and ventral ectoderm markers. 145

**Fig. 3.2.10** *hSMOC1* mRNA cannot rescue bmp2mo bmp2bMO morphants. 146

**Fig. 3.2.11** Zebrafish embryos embedded in 2% agarose survived overnight incubation if the left is in chorions. 147

**Fig. 3.2.12** H2A gfp zebrafish embryos we used to optimize GFP imaging of live zebrafish embryos. 148

**Fig. 3.2.13** Incubation of embryos in 2% low melting point agarose was invariably lethal if it they were dechorionated at the sphere stage of development. 149

**Fig. 3.2.14** The size of the optic cup is unaffected in *zsmoc1* morphants in rx3:gfp reporter lines. 152

**Fig. 3.2.15** Upregulation of BMP signalling in the dorsal retinal of zsmoc-1 morphants. 155

**Fig. 3.2.16** *zsmoc1* constrains the sagittal p-smad gradient zebrafish at 60% epiboly in zebrafish. 159

**Fig. 3.3.1** Suggested model of zsmoc-1 action. 167

**Fig. 3.3.2** Suggested model of zsmoc-1 action - in embryos where bmp signalling has been up regulated with bmp2b overexpression. 168

**Fig. 3.3.3** Suggested model of zsmoc-1 action - in embryos where bmp signalling has been ablated with bmp2b MO. 169

**Fig. 3.3.4** Suggested model of zsmoc-1 action - in embryos injected with *hSMOC1*. 170

**Fig. 3.3.5** Suggested model of zsmoc-1 action - in embryos injected with zsmoc1MO under RED. 172

**Fig. 4.1.1** TALE and TALEN proteins, their structure and action. 186

**Fig. 4.1.2** The three main approaches for TALEN synthesis. 191

**Fig. 4.2.1** The TALEN binding sites. 194

**Fig. 4.2.2** Example restriction digestions used to confirm the length of RVD. 200

**Fig. 4.2.3** Confirmation of the RVD structure after cloning using Sanger sequencing. 201

**Fig. 4.2.4** The results of the first round of Golden Gate cloning to synthesis 5'Tal1zsmoc1. 202

**Fig. 4.2.5** The results of the first round of Golden Gate cloning to synthesis 3'Tal1zsmoc1. 203

**Fig. 4.2.6** The results of the first round of Golden Gate cloning to synthesis 5'Tal2zsmoc1. 204

**Fig. 4.2.7** The results of the first round of Golden Gate cloning to synthesis 5'TALzsmoc1/2. 205

**Fig. 4.2.8** The second Golden Gate cloning reaction for the synthesis of 5'DD and 3'RRTAL1zsmoc1. 207

**Fig. 4.2.9** The second Golden Gate cloning reaction for the synthesis of 5'DD and 3'RRTAL2zsmoc1. 208

**Fig. 4.2.10** Schematic of in vitro assessment of TALEN pair cleavage strategy using southern blot. 210

**Fig. 4.2.11** 5'DD and 3'RRTalzsmoc1 constructs are fully translated in vitro for both Tal1 and Tal2. 212

**Fig. 4.2.12** Southern blot analysis of the DNA cleavage activity of in vitro. 214

**Fig. 4.2.13** The effect of mRNA concentration on the toxicity of TALEN pair injection. 216

**Fig. 4.2.14** Isolating the target site by PCR. 218

**Fig. 4.2.15** Ion Torrent Primer sequence designed for zsmoc1. 219

**Fig. 4.2.16** A schematic representation of the mutation detection strategy followed using Ion Torrent sequencing. 221

**Fig. 4.2.17** No novel mutations were found in *zsmoc1* in embryos treated with *zsmoc1* targeting TALEN pairs. 223

**Fig. 4.2.18** Clonally analysis of the cleavage activity of *zsmoc1* showed no indels as a result of TALEN Treatment. 225

**Fig. 4.2.19** Both TALEN constructs are expressed in vivo. 227

**Fig. 4.3.1** A schematic representation of a type II CRISPR system. 234

**Fig. 6.2.1** A diagram depicting the two microinjections methods utilised. 291

**Fig. 6.3.1** A diagrammatic explanation of the embedding of zebrafish embryos. 294

**Fig. 6.3.2** A diagrammatic explanation of the axes used to measure eye field size in live imaging experiments. 296

**Fig. 6.3.3** An example of gradient thresholding. 298

# List of Tables


| Table Title | Page |
|---|---|
| **Table 1.2.1** The transcription factors and signalling pathways known to be a involved in the development of a vertebrate eye. | 47 |
| **Table 1.2.2** The signalling pathways genes linked to anophthalmia/microphthalmia phenotype. | 52 |
| **Table 2.1.1** Summary of the zebrafish and *Xenopus* genome projects. | 74 |
| **Table 3.1.1** The components of the core bmp signalling pathway. | 110 |
| **Table 3.1.2** Known modulators of bmp signalling. | 117 |
| **Table 3.1.3** Summary on the zsmoc1 targeted morpholinos used during project | 129 |
| **Table 6.2.1** details of Morpholino utilised. | 289 |
| **Table 6.4.1** Plasmid purification details. | 300 |
| **Table 6.4.2** cDNA synthesis program. | 304 |
| **Table 6.4.3** Fasttaq PCR program. | 313 |
| **Table 6.4.4** Phusionflash PCR program. | 314 |
| **Table 6.4.5** SYBR green PCR program. | 316 |
| **Table 6.4.6** RACE PCR program. | 318 |
| **Table 6.5.1** TALEN synthesis program 1 | 335 |
| **Table 6.5.2** TALEN synthesis program 2 | 335 |

# Abbreviations

**General science**

| | | |
|---|---|---|
| A-P | : | anterior posterior |
| D-V | : | Dorsal ventral |
| bp | : | base pair |
| DNA | : | deoxyribonucleic acid |
| cDNA | : | single-stranded complementary DNA |
| gDNA | : | genomic DNA |
| RNA | : | ribonucleic acid |
| ECM | : | extracellular matrix |
| EF | : | eye field |
| LP | : | lens placode |
| LV | : | lens vesicle |
| mRNA | : | messenger ribonucleic acid |
| NMD | : | nonsense mediated decay |
| NR | : | neural retina |
| nt | : | nucleotide |
| OAS | : | Ophthalmo–acromelic syndrome (aka Waardenburg anophthalmia syndrome) |
| AP | : | Animal pole |
| VP | : | Vegital pole |

| | | |
|---|---|---|
| RPE | : | retinal pigment epithelium |
| NR | : | Neural retina |
| OV | : | optic vesicle |
| OC | : | optic cup |
| OF | : | optic fissure |
| OPT | : | Optical projection tomography |
| ddH20 | : | Double Distilled Water |
| RLM-RACE | : | RNA Ligase Mediated Rapid Amplification of cDNA Ends |
| Amp | : | Ampicillin |
| Kan | : | Kanamycin |
| Spec | : | Spectinomycin |
| Tet | : | Tetracycline |
| *ccdB* | : | ccdB toxin. Lethal to cells when not complimented by *ccda* |

**Units of measurement**

| | | |
|---|---|---|
| L | : | litre |
| ml | : | millilitres |
| µl | : | microlitre |
| M | : | molar |
| mM | : | millimolar |
| µM | : | micromole |

| | | |
|---|---|---|
| m | : | metre |
| cm | : | centimetre |
| mm | : | millimetre |
| µm | : | micro-meter |
| bp | : | base pair |
| kb | : | Kilobase |
| Mb | : | Megabase |
| kDa | : | kilodaltons |
| g | : | grams |
| mg | : | milligrams |
| µg | : | micrograms |
| ng | : | nanograms |
| pg | : | picogram |
| mg/µl | : | milligrams per microlitre |
| µg/µl | : | micrograms per microlitre |
| ng/µl | : | nanograms per microlitre |
| mg/µl | : | milligrams per microlitre |
| µg/µl | : | micrograms per microlitre |
| ng/µl | : | nanograms per microlitre |
| pH | : | decimal logarithmic activity of dissolved hydrogen ions |
| °C | : | degree Celsius |

g : x gravity

h : hours

s : seconds

rpm : revolutions per minute

V : volts

mA : milliamperes

Ct : critical threshold

*n* : number

min : minutes

h : hours

**Reagents**

BSA : Bovine serum albumen

DEPC : Diethyl dicarbonate

$ddH_2O$ : Distilled deionised water

DMSO : Di Methyl Sulfoxide

EDTA : Ethylenediaminetetraacetic acid

EtOH : Ethanol

FCS : Foetal calf serum

HEPES : 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

MAb : Monoclonal antibody

PAb : Polyclonal antibody

PBS : Phosphate buffered saline

| | | |
|---|---|---|
| PBST | : | PBS Tween-20 |
| PCR | : | Polymerases chain reaction |
| PFA | : | Paraformaldehyde |
| TBE | : | Tris/borate/EDTA |
| TET | : | Tetracycline |
| E3 | : | Embryo growth medium |
| KCL | : | Potassium chloride |
| KOH | : | Potassium hydroxide |
| $H_2O_2$ | : | hydrogen peroxide |
| LB Broth | : | Lysogeny Broth |
| MeOH | : | Methanol |
| NaCl | : | Sodium chloride |
| NaOH | : | sodium hydroxide |
| NP40 | : | Tergitol-type NP-40, which is nonyl phenoxypolyethoxylethanol |
| RIPA Buffer | : | Radioimmunoprecipitation assay buffer |
| SDS | : | Sodium dodecyl sulphate buffer |
| SSC | : | Saline-sodium citrate buffer |
| dNTP | : | deoxynucleotide triphosphate |
| MOPs | : | 3-(N-morpholino)propanesulfonic acid |
| MES | : | 2-(N-morpholino)ethanesulfonic acid |
| IPTG | : | Isopropyl thiogalactoside, |

| | | |
|---|---|---|
| Xgal | : | 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside |

**Genes**

| | | |
|---|---|---|
| GFP | : | Green fluorescent protein |
| *SMOC1* | : | Secreted modular calcium-binding protein 1 (gene) |
| *SMOC2* | : | Secreted modular calcium-binding protein 2 (gene) |
| SMOC-1 | : | Secreted modular calcium-binding protein 1 (protein) |
| SMOC-2 | : | Secreted modular calcium-binding protein 2 (protein) |
| BMP | : | bone morphogenetic protein |
| Dpp | : | Decapentaplegic |
| MAD | : | Mothers against Decapentaplegic |
| smad | : | Mothers against Decapentaplegic homologue |
| p-smad | : | phosphorylated smad |
| FGF | : | fibroblast growth factor |
| TGF | : | transforming growth factor beta |
| Shh | : | Sonic hedghog |
| tm | : | melting temperature |
| MO | : | morpholino oligonucleotide |
| TAL effector | : | transcription activator-like effector |
| TALEN | : | transcription activator-like effector Nuclease |

| | | |
|---|---|---|
| ZFN | : | Zinc finger nuclease |
| CRISPR | : | clustered regularly interspaced short palindromic repeats |

**Practical**

| | | |
|---|---|---|
| Hpf | : | hour post fertilisation |
| Dpf | : | days post fertilisation |
| O/N | : | overnight |
| ISH | : | in situ hybridisation |
| SDS-PAGE | : | Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophores |
| WB | : | Western blotting |
| SB | : | southern blot |
| PCR | : | polymerases chain reaction |
| RT-PCR | : | reverse transcription- PCR |
| RACE | : | Rapid Application of cDNA Ends |
| q-PCR | : | quantified - PCR or real time PCR |
| BLAT | : | Basic local alignment tool |
| NCBI | : | National Centre for Biotechnology Information |
| OMIM | : | Online Mendelian Inheritance in Man |
| UCSC | : | University of California Santa Cruz |
| PN | : | Product NO |

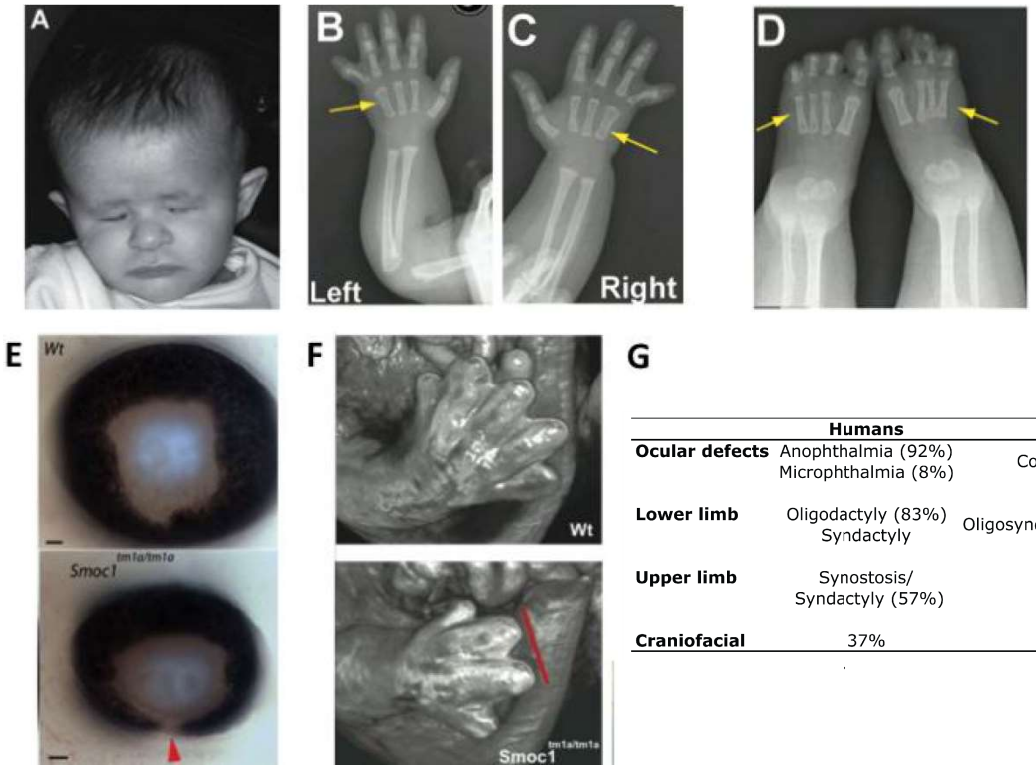
# Chapter 1

## Introduction

# 1.1 Ophthalmo-acromelic syndrome

## 1.1.1 Clinical features

Ophthalmo-acromelic syndrome (OAS or Waardenburg anophthalmia) is caused by mutations in *SPARC related modular calcium binding 1 (SMOC1)* and is characterised by a spectrum of eye phenotypes, with eye malformations ranging from microphthalmia to true anophthalmia (Tekin et al. 2000). However a significant proportion of the known cases (92%) have bilateral anophthalmia **(Fig 1.1.1 A;** Tekin 2002; Tekin et al. 2000; Cogulu et al. 2000; Kara et al. 2002; Garavelli et al. 2006). True anophthalmia is the most severe eye malformation, describing complete absence of the eye (no globe, optic tract, chiasm or tract), and is taken to indicate a catastrophic failure of early eye development (Fitzpatrick & van Heyningen 2005). Other gene mutations which cause anophthalmia in humans have been found in *PAX6* (Glaser et al. 1994), *SOX2* (Fantes et al. 2003), *OTX2* (Ragge et al. 2005), *STRA6* (Pasutto et al. 2007) and *BMP4* (Bakrania et al. 2008), all of which are crucial to early mammalian eye development.




| | Humans | Mice |
|---|---|---|
| **Ocular defects** | Anophthalmia (92%)<br>Microphthalmia (8%) | Coloboma |
| **Lower limb** | Oligodactyly (83%)<br>Syndactyly | Oligosyndactyly (83%) |
| **Upper limb** | Synostosis/<br>Syndactyly (57%) | N/A |
| **Craniofacial** | 37% | 27% |

**Fig 1.1.1 Ophthalmo-acromelic syndrome and *Smoc1* tm1a/tm1a mice are both characterised by bilateral eye and limb malformations.**

A patient showing the typical features of OAS (A) anophthalmia (B and C) fused 4/5th metacarpals in the hands and (D) oligodactyly in the feet. Diagram taken from Garavali et al. (2006). *Smoc1*$^{tm1a/tm1a}$ Mice show an overlapping phenotype with OAS, most notably coloboma in the eye (E) and oligosyndactyly in the lower limbs (F). This data is summarised in (G) Rainger et al. (2011)

---

The other defining feature of OAS is a distinctive pattern of limb malformations with the lower limbs showing the most severe defects. Postaxial oligodactyly is observed in 83% of patients, **(Fig. 1.1.1 D)** and with the remaining cases exhibiting some form of osseous or soft tissue syndactyly of the 4$^{th}$ and 5$^{th}$ ray (Tekin et al. 2000; Cogulu et al. 2000; Kara et al. 2002; Garavelli et al. 2006). The upper limb malformations are less severe, ranging from soft tissue syndactyly (57%; **Fig. 1.1.1 C)** to a shortening of the 5$^{th}$ digit. Other phenotypes associated

with OAS include mental retardation (37%) and craniofacial abnormalities (37%; **Fig. 1.1.1 G;** Rainger et al. 2011).

## 1.1.2 The identification of mutations in *SMOC1*, the principal genetic cause of OAS

Three separate research groups have identified homozygous mutations of *SMOC1* as the cause of OAS (Rainger et al. 2011; Abouzeid et al. 2011; Okada et al. 2011).

Our group identified eight separate mutations in eight families. Six of these are loss of function mutations as they are either nonsense or frameshift mutations (Rainger et al. 2011). The last two missense mutations are predicted to be loss of function, as they completely phenocopy the nonsense mutations. Additionally, the amino acid residues that are changed in the missense mutations are evolutionarily well conserved; the Arg278 residue is conserved to *Xenopus laevis* and the Thr283 residue is conserved to *Drosophila melanogaster* (Rainger et al. 2011). Of the six families where a *SMOC1* mutation has not been identified, two mutations map to the 14q24 locus and are likely to be as yet unidentified *SMOC1* regulatory mutations. This suggests that *SMOC1* mutations account for ~70% of OAS cases (Rainger et al. 2011). Independent research from two other research groups has identified four different mutations in *SMOC1,* all of which are predicted to be loss of function mutations (either nonsense or splice site mutations; Okada et al. 2011; Abouzeid et al. 2011).

## 1.1.3 The structure of SMOC-1

SMOC-1 is a member of the BM40 family of proteins and along with its close homologue, SMOC-2, it is made up of six domains: one signal peptide and one Follistatin-like domain, two Thyroglobulin-like domains, a unique SMOC domain and two EF hand domains (Fig 1.1.2). (Vannahme et al. 2002; Vannahme et al. 2003). The protein also contains a signal peptide sequence. The signal peptide sequence is a short string of 5-30 hydrophobic amino acids followed by a signal peptidase cleavage site. The cleavage of signal peptides targets the protein to the secretory pathway. As the SMOC proteins lack the long strings of hydrophobic amino acids that would be necessary for a transmembrane protein this structure would predict that the protein is secreted and not a transmembrane protein. This result is borne out by experimental data with Smoc-1 and Smoc-2 which are both secreted into the media in cell culture. Whole mount immunohistochemistry of both the testes and the ovaries has shown that SMOC-1 is localised to the basement membrane in adult mice (Vannahme et al. 2002; Vannahme et al. 2003).

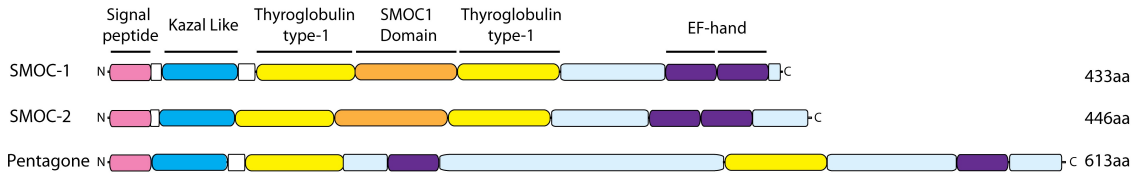



**Fig 1.1.2 A side by side comparison of the structure of theSMOC subfamily of BM40 proteins and related *Drosophila* Pentagone.**

Both SMOC-1 and 2 are based on the human amino acid sequence (UniProt: Q9H4F8 and Q9H3U7), while the Pentagone sequence is based on the *Drosophila* amino acid sequence reported in Vuilleumier et al. 2010

Once the signal peptide has been cleaved off, the mature SMOC peptide starts with a Kazal like domain followed by two thyroglobulin-like domains. These related domains in other proteins have been shown to inhibit protein degradation by blocking peptidase activity. Notably, the second thyroglobulin domain appears to be critical for SMOC-1 as this domain harbors the only two missense mutations identified (Rainger et al. 2011).

The thyroglobulin domains flank 65 conserved amino acids, which have been called the SMOC domain. However there is no compelling evidence of a specific function attached to the SMOC domain, and it is possible that it is made up of several smaller structural motifs each with a distinctive function (Vannahme et al. 2002; Vannahme et al. 2003).

The final functional groups are a pair of EF hand domains containing glycosaminoglycan-binding motifs. These motifs allow SMOC proteins to bind to heparin and heparin sulphate and it has been shown that in cell culture SMOC-1's adhesion to the cell membrane is dependent on the presence of heparin sulphate proteins (e.g. the glypicans). If the glycosaminoglycan-binding motifs are mutated then the protein will no longer adhere to the cell membrane (Klemenčič et al. 2013).

## 1.1.4 The mode of action of Smoc-1 in animal models

While Smoc-1 is not a well studied protein, the developmental data from animal models supports the human genetic assignment of this as the causative gene in OAS.

Two separate papers detailing the phenotypes of gene trap mice have been published, and both have remarkably similar features (*Smoc1*$^{tm1a/tm1a}$ (Rainger et al. 2011) and *Smoc1*$^{Tp/Tp}$ (Okada et al. 2011)). A compelling case can be made that they phenocopy OAS in humans. As in OAS cases both show more severe defects in the lower limbs than upper limbs. The hind limbs displayed both oligodactyly and syndactyly, while the forelimbs displayed no phenotype. The *Smoc1*$^{tm1a/tm1a}$ mice had a significant percentage of embryos that showed a cleft palate phenotype (27.3%), which is also observed in OAS (Rainger et al. 2011). This is consistent with the presence of expression patterns observed in both *in situ* hybridisations of E9.5 and E10.5 mouse embryos which show distinctive expression patterns in both the first branchial arch and the limb buds (Rainger et al. 2011; **Fig 1.1.3 a & b**).

The most surprising aspect of the mouse phenotype is that the eye defects are comparatively mild when compared to the true anophthalmia displayed in OAS patients. The *Smoc1*$^{tm1a/tm1a}$ mice show a mild coloboma phenotype while *Smoc1*$^{Tp/Tp}$ mice show mild microphthalmia and mild defects in closure of the optic fissure. It is thought this might be due to leaky expression from the gene trap (*Smoc1*$^{tm1a/tm1a}$ had residual expression of 10% of the *Smoc1* transcript). However,

a complete null has now been generated, and the ocular phenotype remained mild. (Rainger personal correspondence 2012). It remains possible

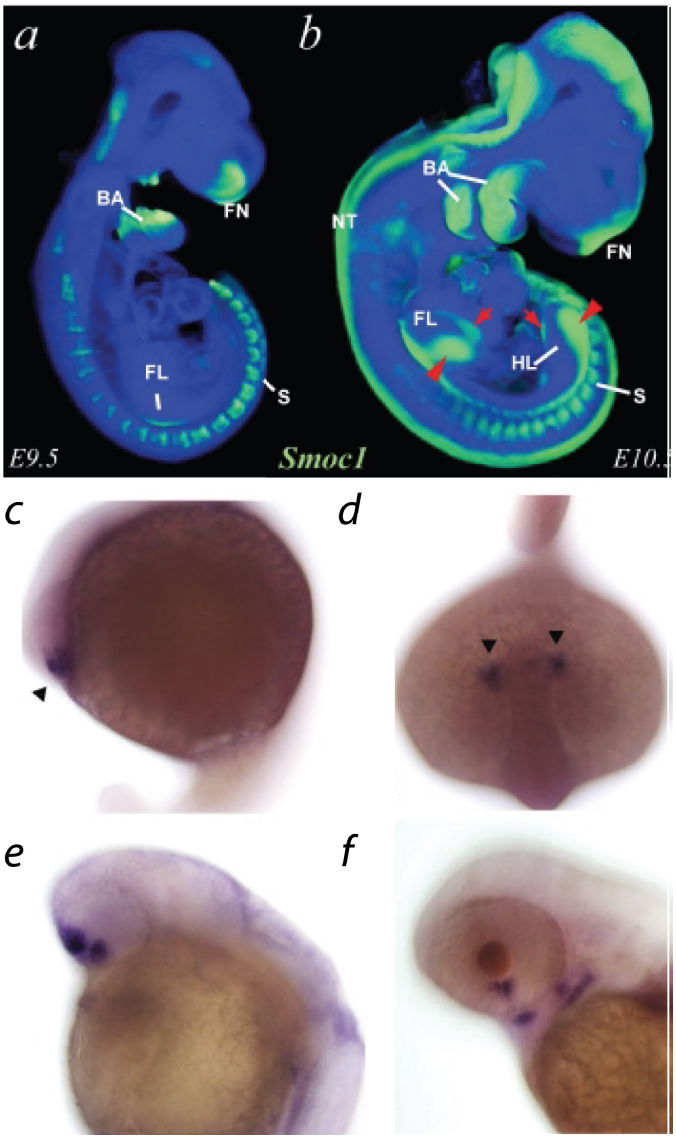



**Fig 1.1.3 The known expression patterns of *SMOC-1* during development.**

**(A and B)** OPT models of whole mount *in situ* at both E9.5 and E10.5 respectively. **(A)** At E9.5 *SMOC1* expression is seen in the branchial arches (BA), the fronto-nasal region (FN), the forelimbs (FL) and the somites (S). **(B).**By E.10.5 expression has expanded to the notochord. In the forward and hind limbs expression is confined medially in both to the dorsal and ventral axes..
**(C, D, E and F if a)** Whole mount RNA *in situ* of zebrafish showing that *smoc1* expression is initially localised within the anterior of the RPE by 18 SS (~16 hpf) **(C and D).** By 24 hpf expression is clearly compliant to the ventral retina flanking the optic fissure **(E).** By 48 hpf *smoc1* expression can also be seen in the branchial arches.

Figures taken from (Rainger et al. 2011) and (Abouzeid et al. 2011)

---

that *smoc-1* is not a significant player in the development of the mouse eye. It is expressed at much lower levels in the eye than either the branchial arch or limb buds. Rainger et. al failed to detect *Smoc-1* transcripts in the eye using *in situ* hybridisation. Yet the *Smoc1*$^{tm1a/tm1a}$ LacZ reporter was active in the eye. This is

consistent with the observation that *Smoc-1* down regulation causes a mild eye phenotype (Rainger et al. 2011).

The only other significant functional work published in vertebrates is in *Xenopus*. Unilateral *xsmoc1* depletion with morpholino oligonucleotides leads *Xenopus* embryos to develop unilateral anophthalmia on the side of the injection (Thomas et al. 2009). However by 2011 Abouzeid et al. 2011 had demonstrated that *smoc1* is expressed in the zebrafish eye by 16 hpf and reported that *smoc1* depletion results in a coloboma phenotype suggesting that the zebrafish could be a good model organism for the study of OAS.

Outside of vertebrates some compelling work has been done in *Drosophila*. Mutations in the *Drosophila smoc1* orthologue *Pentagone* result in the loss of lateral identity, exemplified by the absence of the fifth longitudinal vein (Vuilleumier et al. 2010). While the structure is not homologous to the vertebrate limb, it does have obvious similarities to the oligodactyly observed in both humans and mice (Vuilleumier et al. 2010). The imaginal wing discs are a well-established model system for bmp signalling. It is possible to dynamically monitor the dpp (the *Drosophila* homologue of BMP) gradient in real time and is relatively easy to manipulate the other components of the BMP signalling pathway. In fact, the *pentagone* expressed in the imaginal wing discs allowed Vuilleumier et al. to investigate its effect on the *dpp* gradient. This led them to propose a molecular model of Pentagone action**(Fig 1.1.4)**

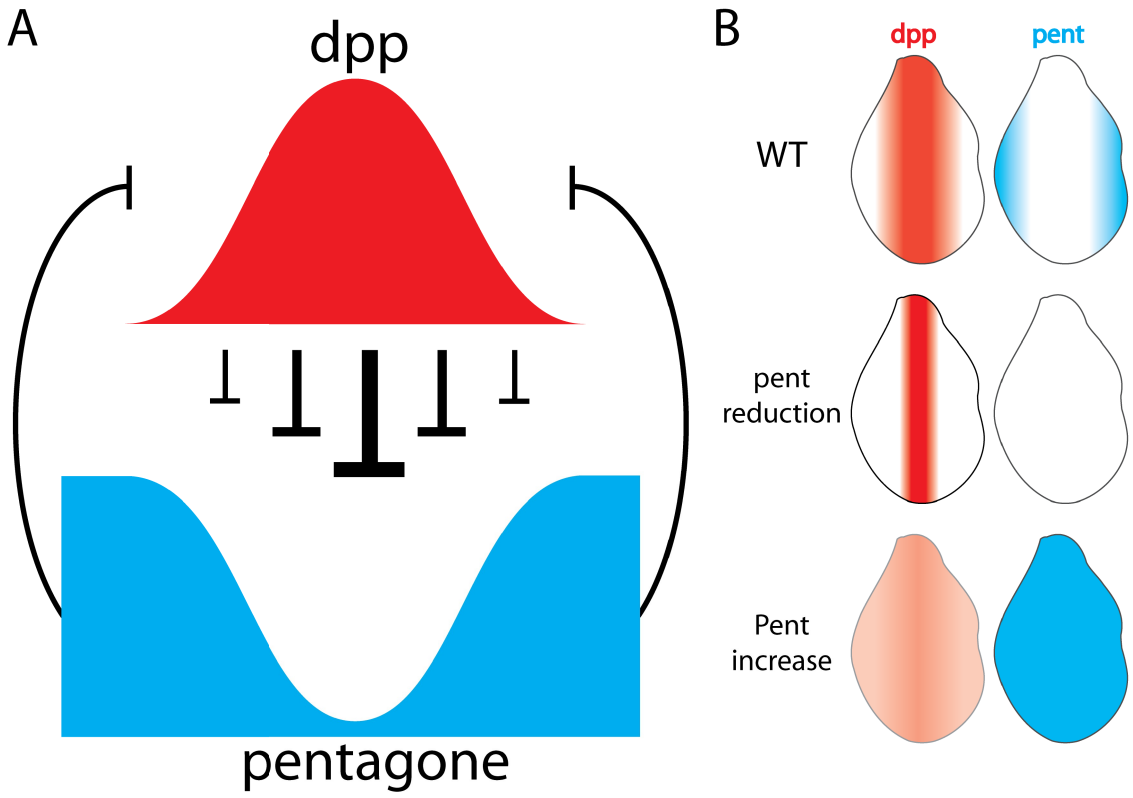



**Fig 1.1.4 The Vuilleumier model of Pentagone action in *Drosophila***

**(A)** Under the Vuilleumier model dpp represses *pentagone* expression. This excludes pentagone from the region of dpp activity which leads to the establisment of a pentagone countergradient to dpp. Pentagone inturn inhibits dpp activity by blocking the endocytosis of dpp. **(B)** Schematic representaion of the effect of dpp and pentagone levels in the *Drosophila* imaginal wing discs, In wildtype (WT) flies the cenral band of dpp expression restricts pentagone expression to the lateral wingdisk, forming a feedback loop which fine-tunes the Dpp gradient. If pentagone is reduced (as in the case of the pent mutant embryos) the dpp is endocytosed near source. Resulting in a narrower but more intense band of dpp activity. If pentagone is upregulated (if pentagone is expressed with the dpp promoter) then endocytosis is blocked. Resulting in a broader but less intense band of dpp activity.

Figures derived with information taken from Vuilleumier et al. 2010

---

*Pentagone (pent)* was initially identified as the gene regulated by BMP responsive silencer elements. This was later confirmed genetically in the original this; not only was *pentagone* expression confined to the lateral edges of the imaginal disc away from the midline where dpp levels were greatest but in flies

where *dpp* was ectopically expressed throughout the wing disc pentagon expression was eliminated. They went on to show that pentagone itself altered dpp signalling. In *Pent* mutants the region of dpp signalling was contracted though the intensity of mad phosphorylation (and thus dpp signalling) was significantly increased. Conversely, where Pentagon was ectopically expressed the gradient is broader but mad? phosphorylation was significantly reduced. They went on to show that *pent* mutants could be rescued by down regulating endocytosis. This lead them to propose a model where pentagone acted not as a simple BMP antagonist but as a gradient expander. In this capacity it acts in a feedback loop with dpp. Hamaratoglu et al. went on to expand upon this work and showed that this feedback loop was critical in controlling the shape of the *Drosophila* wing as it grows in size (Hamaratoglu et al. 2011).

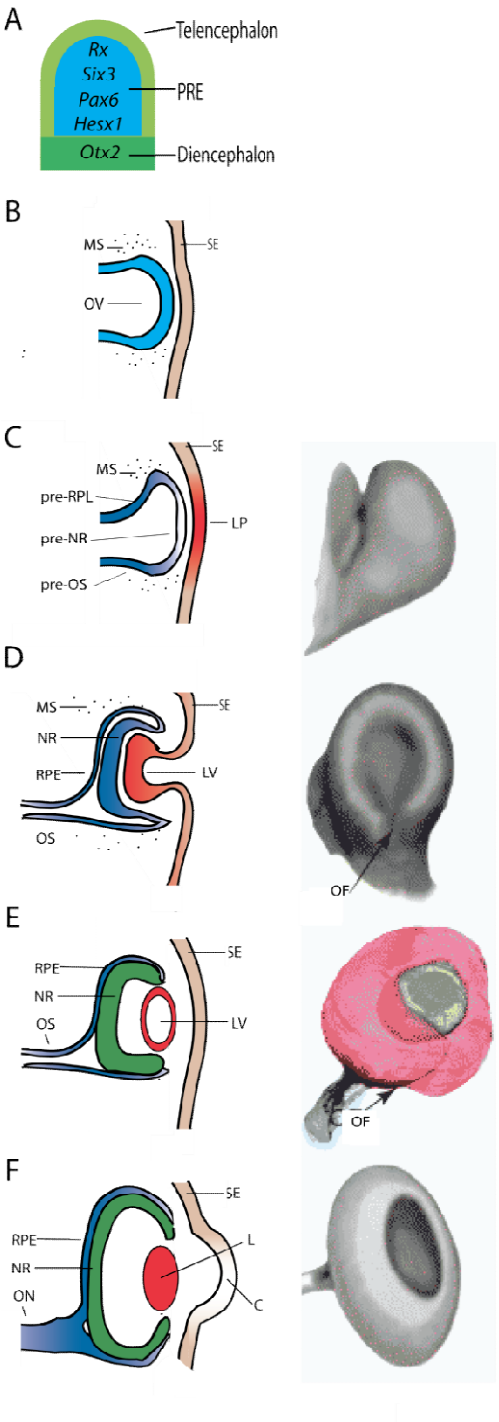
# 1.2 Eye development

## 1.2.1 Vertebrate eye development

Vertebrate eye development can be broadly divided into five separate stages.

1. Specification of the eye field **(Fig. 1.2.1 A).**
2. Formation of the optic vesicle (OV) **(Fig. 1.2.1 B).**
3. Lens(L) induction **(Fig. 1.2.1 C)**
4. Organization of the neural retina and retinal pigment epithelium **(Fig. 1.2.1 D-E)**
5. Fusion of the optic fissure **(Fig. 1.2.1 E-F**; reviewed in Fitzpatrick & van Heyningen 2005; Adler & Canto-Soler 2007; Chow & Lang 2001).

It should be noted that lens induction, fusion of optic fissure and the early stages of the organisation of the neural retina occur simultaneously **(Fig. 1.2.2).**




| | |
|---|---|
| Human | - GD |
| Mouse | 8.5 GD |
| Chicken | 0.9 GD |
| Zebrafish | 10 hpf |

| | |
|---|---|
| Human | 28 GD |
| Mouse | 9 GD |
| Chicken | 1.3 GD |
| Zebrafish | 10.5 hpf |

| | |
|---|---|
| Human | 32 GD |
| Mouse | 9.5 GD |
| Chicken | 2 GD |
| Zebrafish | 11 hpf |

| | |
|---|---|
| Human | 33-37 GD |
| Mouse | 10 GD |
| Chicken | 2.2 GD |
| Zebrafish | 26 hpf |

| | |
|---|---|
| Human | 41 GD |
| Mouse | 12 GD |
| Chicken | 2.8 GD |
| Zebrafish | 26 hpf |

| | |
|---|---|
| Human | >56 GD |
| Mouse | 16.5 GD |
| Chicken | 7 GD |
| Zebrafish | 48 hpf |

**PRE:** presumptive retinal ectoderm **D:** dorsal**; V:** ventral**; MS:** mesenchyme; **SE:** surface ectoderm; **OV:** optic vesicle; **LP:** lens placode; **pre-:** presumptive; **RPE:** retinal pigment epithelium; **NR:** neural retina; **LV:** lens vesicle; **OS:** optic stalk; **OF:** optic fissure; **ON:** optic nerve; **hpf:** Hours post fertilisation; **GD:** gestation day. **C:** cornea **L:** lens

**Fig. 1.2.1 Schematic of the development of a generic vertebrate eye.**

**(A)** The presumptive retinal endoderm (PRE) is specified in the anterior neural plate. The transcription factors known to be involved are shown. **(B)** The diencephalon envaginates towards the SE. **(C)** Once the diencephalon makes contact with the surface ectoderm (SE) it differentiates into the presumptive retinal pigment epithelium (RPE), neural retina (NR) and optic stalk (OS). **(D)** This in turn induces the SE to form the lens placode (LP). This is followed by the invagination of the optic vesicle. At this stage the OF begins to form in the ventral eyecup**.** As the optic cup matures the lens vesicle (LV) separates from the SE before becoming the solid lens(L). The lens induces the SE to differentiate into the C. The gap between the NR and RPE closes and the NR fully differentiates**(E-F)**.
Meanwhile the optic stalk forms the passage for the optic nerve**(F)**. Finally the optic fissure closes **(F)**.

Figure based on information taken from Adler & Canto-Soler 2007; Kimmel et al. 1995, Glass & Dahm 2004; Ohuchi et al. 1999; Mathers et al. 1997 and Fitzpatrick & van Heyningen 2005

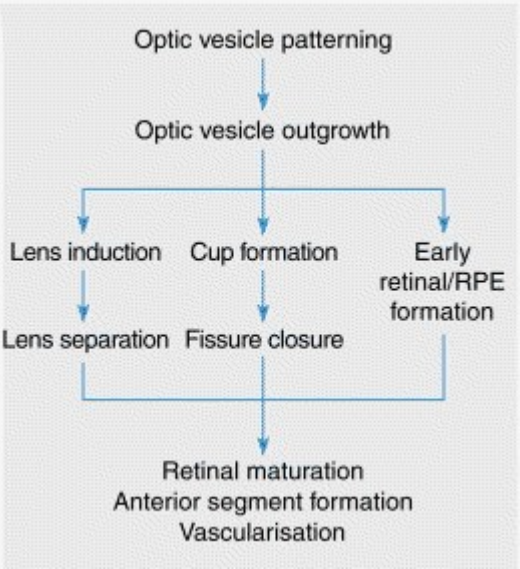



**Fig. 1.2.2 the sequence of development for the vertebrate eye.**

Schematic diagram showing main processes involved in eye development and the relative order which they occur in.
Taken from Fitzpatrick & van Heyningen 2005

For over 50 years it has been known that eye field formation is dependent on neural induction. Nieuwkoop demonstrated this by explanting frog neural-ectoderm which differentiated into optic tissue if the cells of the explant were first disassociated (Nieuwkoop 1963). This differentiation occurred even in the absence of mesodermal tissue, showing that the development is cell autonomous (Nieuwkoop 1963; Chow & Lang 2001). It has since been established that the vertebrate eye is derived from the pre-lens placode. After the neural tube has closed the anterior neural plate is potentiated for eye development by the expression of noggin and otx2 (Zuber et al. 2003; Chuang & Raymond 2002). The presumptive retinal ectoderm (PRE) is derived from this region of the anterior neural plate (Donner et al. 2006;Streit 2007; Lachke & Maas 2010; Varga et al. 1999). Surrounded by the pre-telencephalon and pre-diencephalon (Fernández-Garre et al. 2002; Inoue et al. 2000), the PRE is defined by the positive regulation of Fgf signalling and simultaneous inhibition by Bmp and Wnt signalling **(Fig. 1.2.1 A**; Lachke & Maas 2010**;** Macdonald et al. 1995). This leads to the up-regulation of multiple transcription factors, most notably Lhx2 Rx, Six3, Pax6, Hesx1 and Six6 (Lachke & Maas 2010; Zuber et al. 2003).

The correct establishment of the eye field requires that these transcription factors act simultaneously and cooperatively. Knocking out any one of these genes results in significant eye defects (Zhang et al. 2000; Mathers et al. 1997; Roy et al. 2013; Dattani et al. 1998; Wallis et al. 1999). Yet, no transcription factor fully controls the expression of any of the others. *Rx* is the earliest marker of the eye field yet does not initiate *Pax6*, *Six3*, or *Otx2* expression within the neural plate. This was demonstrated by Zhang *et al.* (2000) with knockout mice. The presence of Rx is required to up regulate the expression of each of these genes during the specification of the eye field (Zhang et al. 2000). Conversely *Pax6* null mice still initiate *Rx*, *Six3*, and *Otx2* expression (Chow & Lang 2001).

While these results appear contradictory they can be explained by either common external signals and/or via cross regulation. The role of Wnt, Fgf and Bmp signalling during eye development is already well-established (Lachke & Maas 2010) and it appears that cross regulation is also a factor. This was established by the injection of various "cocktails" of different combinations of eye field transcription factors (*Rx, Six3, Pax6*, *tll* and *Optx2*) into a one cell *Xenopus* embryo, the resulting overexpression leads to the formation of ectopic eyes in 88% of *Xenopus embryos* (Zuber et al. 2003). This demonstrates that these transcription factors are sufficient to drive eye field specification. Zuber *et al.* went on to show that removing any one gene from the mRNA cocktail reduced but did not eliminate the formation of ectopic eyes (between 24% for a mRNA cocktail minus Pax6 and 72% for a mRNA cocktail missing *Optx2*; Zuber et al. 2003).

Within the PRE the cell fate is restricted along the midline of the PRE until it is separated into two distinct eye fields by Shh signalling (Chiang et al. 1996; Macdonald et al. 1995). When these separate eye fields have been established the optic vesicle (OV) begins to form by bilaterally envaginating towards the surface ectoderm (SE). Once the OV has made contact with the SE both tissues begin to differentiate. The OV induces the formation of the lens placode (LP). Then the LP in turn induces the OV to differentiate into the presumptive retinal pigment epithelium (RPE), neural retina (NR) and the optic stalk (OS). After the lens placode has been induced it begins to invaginate forming the lens vesicle (LV). Simultaneously OV also invaginates surrounding the (LV) and coincidentally the NR, RPE and OS form.

As the optic cup changes morphology the optic fissure is formed. The RPE and NR begin to fuse approximately in the middle of the longitudinal axis of the developing eye. This fusion proceeds simultaneously proximally and distally, creating a single unified layer of neural ectoderm. Once the fusion front reaches the OS, it closes forming the optic nerve (ON). By the time the lens has induced the formation of the cornea (C) the optic fissure has closed and the optic globe is complete (reviewed in Adler & Canto-Soler 2007; and Fitzpatrick & van Heyningen 2005).

This is the generic model of vertebrate eye development and it should be noted that there is interspecies variation in the development of the eye. For example in teleost fish (i.e. zebrafish and medaka) a pair of OV aren't formed by the invagination of the diencephalon. Instead the eye field is formed inside the

telencephalon: it divides to form two independent eye fields, which envaginate into the eyes (Woo & Fraser 1995). This leads to several other developmental differences. For example, the RPE and NR are organised by individual cell migration and not envagination (Rembold et al. 2006; Brown et al. 2010) and the Lens in zebrafish delaminates from the OV rather than invaginating inwards with the RPE (Glass & Dahm 2004). Therefore the LV never forms in zebrafish (Glass & Dahm 2004). However, the broad sequence of events remains largely the same. The function of the main eye differentiation genes is conserved and the relative time points are conserved between species **(Fig. 1.2.1)**, indicating that the signalling events are broadly conserved (Adler & Canto-Soler 2007; Donner et al. 2006).

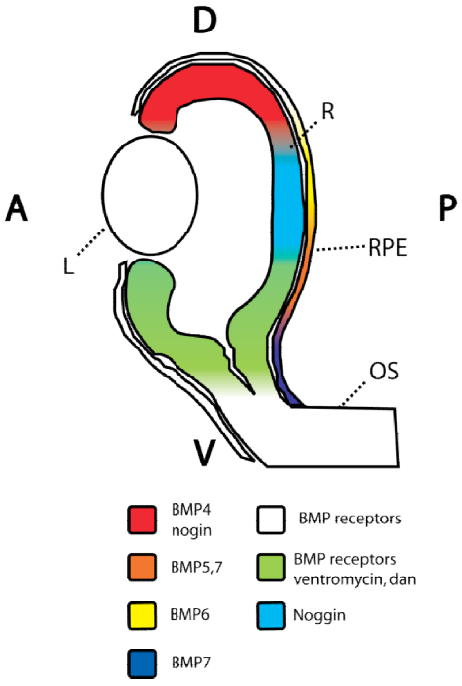



**Fig 1.2.3 Schematic of the expression of Bmp signalling proteins in dorsoventral patterning of the vertebrate eye.**

The figure depicts the regions with highest levels of expression as assessed by in situ hybridisation

OS: optic stalk; R: retina; RPE: retinal pigment epithelium; L: lens
D: dorsal; V: ventral; A: anterior; P: posterior.

Figure based on information taken from
Adler & Canto-Soler 2007.

| Stage of eye development | Transcription factors | Signalling Pathways |
|---|---|---|
| 1.Formation of OV | **Rx; Pax6;** Six3; Lhx2;Six6; ET; tll; Hes1**;Otx2; Sox2** | FGFs; Wnts; **BMPs** |
| 2. L induction | **Rx; Pax6**; tll | Cyclops; SHH**; RA** |
| 3.Patterning of the NR and RPE | **Pax6; Rx;** Lhx2; *Chx10*;**Otx2;** Mitf; Pax2; Foxg1; BF2/Foxd2 | SHH; FGFs; Activin; **BMP;RA** |
| 4. Fusion of the OF | **Pax6;** Lhx2; Hes1 | **RA** |
| 5. Dorsoventral patterning of the NR | **Pax6**; Foxg1;Foxd2; SOHo1; GH6;Pax2; Vax; Tbx5;Xbr1 | Nodal; FGFs; SHH; **BMPs**; **RA ;** Ventroptin; Follistatin; Chordin; Noggin; DAN |

**Table 1.2.1 The transcription factors and signalling pathways known to be involved in the development of a vertebrate eye.**

The table illustrates which transcription factors and signalling cascades are known to be functional in eye development. Genes and signalling pathways in which mutations cause anophthalmia or microphthalmia have been highlighted in bold. Genes only known to cause microphthalmia are underlined.

Information taken from Adler & Canto-Soler 2007 and Williamson & FitzPatrick 2014

As the eye is formed it is simultaneously patterned along several axes. The eye varies both structurally and molecularly in the dorsoventral, anterior posterior and naso-temporal axes. To achieve this several interacting signalling pathways are utilised **(Table 1.2.1)** most notably Wnt signalling, Mapk signalling and Bmp signalling (Adler & Canto-Soler 2007).

An example of this is Bmp signalling in the patterning of the dorsoventral axis. The retina varies morphologically along the dorsoventral axis in a number of ways. There are differential rates of retina expansion, with dorsal tissue expanding faster than ventral tissue (Koshiba-Takeuchi et al. 2000), leading to the formation of the optic fissure which is located in the ventral tissue (Koshiba-Takeuchi et al. 2000). Bmp ligands are expressed in a spatially restricted manner **(Fig. 1.2.3)**. *Bmp4* for example is expressed dorsally and promotes dorsal

differentiation. In *Xenopus* the up-regulation of *bmp4* expression leads to an expansion of the expression of dorsal marker genes (for example *pax6).* There is a corresponding down-regulation of ventral expressed genes (for example *pax2* the *vax2*; Sasagawa et al. 2002). Bmp antagonists like Noggin and Chordin-like1 (ventropin) promotes ventral identity by blocking bmp signalling (Sakuta et al. 2001). Up-regulation of the proteins leads to an expansion of ventral marker genes (for example *pax2* and *vax2).*

Of course, Bmp signalling is not the only pathway to specify dorsal identity. Wnt signalling also promotes the dorsal retina. While both Shh and RA signalling have been shown to promote ventral identity(Sakuta et al. 2001; Marsh-Armstrong et al. 1994).

Eye development is thus dependent on a complex cascade of molecular events. The process is finely balanced, with many processes occurring simultaneously and with a considerable amount of crosstalk. The failure of any one part will impact in the whole process of eye development,which would then lead to catastrophic failure and eye malformations.

## 1.2.2 Known genetic causes of anophthalmia and microphthalmia

The majority of genes known to cause anophthalmia and microphthalmia can be split into two categories encoding either proteins acting on signalling pathways or transcription factors (Adler & Canto-Soler 2007; Williamson & FitzPatrick 2014). This is unsurprising, given that anophthalmia and microphthalmia are developmental disorders. Anophthalmia, as the most severe eye malformation may be reasonably hypothesised to result from mutations in genes acting at the very earliest stages of eye development thus leading to optic vesicle aplasia with consequent complete failure in the initiation of eye development (S Olitsky and L Nelson 2006). Most of the causative genes are expressed during the early stages of eye development. **(Table 1.2.1).** However, in many cases the absence of the eye is the consequence of regression, after some initial development in the optic cup. This appears particularly likely in the cases that are not true anophthalmia as the presence of optic nerve, chiasm and optic tracks suggests regression (Fitzpatrick & van Heyningen 2005). This has been observed in mouse models of anophthalmia. For example *bmp7* knockout mice were shown to initially undergo normal lens and retinal development before degradation. In total 60% of the *bmp7* null mice went on to develop anophthalmia (Dudley et al. 1995). Furthermore, the known causative genes discussed above are typically involved in multiple stages of eye development **(Table 1.2.1)**.

Biallelic mutations in *PAX6* were the first identified genetic cause of anophthalmia, but such compound heterozygous cases are rare (Glaser et al. 1994). Heterozygous *PAX6* loss-of-function mutations are more common and are associated with eye abnormalities; typically aniridia and Peter's anomaly (Hingorani et al. 2012). *PAX6* is a transcription factor that plays a central role throughout eye development. It is expressed at every stage of eye formation (Macdonald & Wilson 1997) and has been shown to be essential to many different processes in eye development **(Table 1.2.1).** This central role has led to *PAX6* acquiring the label of "the master regulator of eye development" (Halder et al. 1995; Baker 2001).

While *PAX6* may be a master regulator there is considerable evidence that it functions downstream of *RX (or RAX). RX* is the earliest known marker of the optic vesicle and it is continually expressed in both the OV and NR throughout eye development (Bailey et al. 2004). *RX* compound heterozygotes have been associated with severe ocular malformations (either anophthalmia or microphthalmia; Chassaing et al. 2014; Voronina et al. 2004).

The strongest evidence that RX acts upstream of PAX6 comes from mice. Homozygous nulls fail to express *Pax6* (Zhang et al. 2000). Additionally, in *Xenopus, Rx* morphants cannot form OV, leading them to display true anophthalmia. This can be rescued with the injection of both *otx2* and *Pax6* mRNA (Mathers et al. 1997).

The most commonly mutated genes in bilateral anophthalmia and severe microphthalmia are *SOX2* and *OTX2* which together account for at least 60% of cases (Gerth-Kahlert et al. 2013). Both genes code for transcription factors. *OTX2*

has a central role in establishing the brain field (Adler & Canto-Soler 2007), from which the OV initially forms. Unsurprisingly patients with *OTX2* mutations often have neurological defects along with ocular malformations (Ragge et al. 2005). Unusually for a gene known to cause anophthalmia *OTX2* is excluded from the OV, which envaginates from the diencephalon (Adler & Canto-Soler 2007). Later in eye development it begins to express in the OS before becoming restricted to the RPE (Martinez-Morales et al. 2001). *OTX2* is thought to be critical to neural retina development. This is supported by the phenotype of *OTX2* heterozygotes. Patients who do not have anophthalmia invariably display retinal defects (Fitzpatrick & van Heyningen 2005). It has been proposed that the whole spectrum of ocular defects found in *OTX2* mutants are caused by the failure in RPE formation, and that the severity of the defects depends on how early RPE development fails (Fitzpatrick & van Heyningen 2005). This is the case in mice, where the $Otx2^{-/-}$ genotype is embryonic lethal due to a failure of gastrulation (Acampora et al. 1995). As in humans, the severity of the ocular malformations that *Otx2* heterozygotes display is extremely variable. $Otx2^{+/-}$ mice display ocular phenotypes, which range from completely normal (Acampora et al. 1995) to displaying anophthalmia (Matsuo et al. 1995) depending on the genetic background. Martinez-Morales et al. demonstrated that Otx proteins have a critical role to play in RPE formation by crossing $Otx2^{+/-}$ and $Otx1^{-/-}$ mice. This led to the almost complete ablation of the RPE, which has the knock-on effect of disrupting NR differentiation (Martinez-Morales et al. 2001).

| BMPs | | | RA | | |
|---|---|---|---|---|---|
| **BMP4** | Ligand | Monoallelic | **STRA6** | RA synthesis | Biallelic |
| **BMP7** | Ligand | Monoallelic | **ALDH1A3** | RA synthesis | Biallelic |
| **GDF6** | Ligand | Monoallelic | **RARB** | RA synthesis | Monoallelic and Biallelic |
| **GDF3** | Ligand | Monoallelic | | | |

**Table 1.2.2 The signalling pathways genes linked to anophthalmia/microphthalmia phenotype**

Information taken from Williamson & FitzPatrick 2014

SOX2 is a central component of the eye development network in humans. The loss of a single allele is the most common cause of anophthalmia in humans (Williamson & FitzPatrick 2014). Work on *SOX2* in model organisms has shown that it is involved in multiple processes during eye formation **(table 1.2.1).** For example *SOX2* acts as a binding partner for multiple eye development genes. The Sox2-Pax6 complex has been shown to be vital for the correct expression of crystallin during the lens development of chickens (Kamachi et al. 2001) and Sox2-Otx2 heterodimers have been shown to regulate Rx expression (Danno et al. 2008).

The second category of vertebrate eye genes are those involved in signalling pathways. Currently two pathways that have been implicated as causes of anophthalmia in humans. They are the BMP signalling pathway and retinoic acid (RA) signalling pathway **(Table 1.2.2;** Williamson & FitzPatrick 2014).

Of the two, the BMP signalling was the first to be implicated in human anophthalmia/microphthalmia(Williamson & FitzPatrick 2014). Currently there

are four genes that have both been linked to anophthalmia/microphthalmia and are components of the BMP signalling pathway.

The first to be discovered was *BMP4* (Bakrania et al. 2008). There have been at least six cases of patients with mono allelic *BMP4* mutations displaying anophthalmia / microphthalmia that have been published(Bakrania et al. 2008; Reis et al. 2011; Zhang et al. 2009). However, there may be other causative genes or confounding environmental effects in this instance as there are two cases where large deletions that include multiple genes, along with *BMP4* have led to milder eye phenotypes (e.g. Rieger anomaly, micro-cornea, nystagmus and glaucoma; Reis et al. 2011). It is clear that *Bmp4* does have a functional role in vertebrate eye development. The expression pattern of *Bmp4* is conserved from mice to humans (Reis et al. 2011) and the *Bmp4* null mice failed to develop the lens placode (Furuta & Hogan 1998; Williamson & FitzPatrick 2014). This appears to be an entirely separate process from OV specification as the expression of *Pax6* remains unaffected (Furuta & Hogan 1998). Of particular note in relation to the symptoms of OAS is that BMP4 heterozygotes also display digit abnormalities, with two unrelated cases exhibiting postaxial polysyndactyly (Williamson & FitzPatrick 2014).

BMP7 is another BMP ligand linked to eye development. Three unrelated cases have been reported of *BMP7* heterozygotes that display anophthalmia/ microphthalmia (Wyatt et al. 2010). If *BMP7* mutations are indeed a cause of anophthalmia/ microphthalmia then either the genetic background or an environmental variable must also play a role because all the causative mutations identified have been inherited from the mother, none of which displayed an eye

phenotype(Wyatt et al. 2010). However, this is in line with the results obtained for *bmp7* knockout mice, which displayed a high variability in phenotype severity depending on the genetic background of the mice used (Dudley et al. 1995; Wawersik et al. 1999). In C3H/He mice 91% of the homozygous nulls developed anophthalmia (Wawersik et al. 1999). Whilst only 50-60% of 129/Sv mice developed anophthalmia(Dudley et al. 1995; Wawersik et al. 1999) and 20% of C2 mice displayed no eye defects at all.

Finally the BMP ligands GDF3 and GDF6 have both been linked to anophthalmia/ microphthalmia (MIM: 601147; Asai-Coakwell et al. 2007; MIM: 606522; Ye et al. 2010). Yet there is still uncertainty as to whether these mutations are causative (Williamson & FitzPatrick 2014). This is because the variants have only been identified in less than 2% of cases of anophthalmia/microphthalmia and every variant currently reported is a heterozygous non-synonymous mutation (Williamson & FitzPatrick 2014) with these variants either inherited from an unaffected parent, or the mode of inheritance was never established (Williamson & FitzPatrick 2014). While GDF6 has been shown to be required for the correct development of the eye in multiple model organisms, it appears to be involved later in eye development. *GDF6* mouse knockouts exhibit abnormal neural retina and display altered electroretinograms (Asai-Coakwell et al. 2013). In addition *GDF6* zebrafish morphants display ventralised neural retina (French et al. 2009; Gosse & Baier 2009) and *Xenopus* morphants showed a reduction in neuronal differentiation (Hanel & Hensey 2006).

More recently the retinoic acid (RA) signalling pathway has been linked to cause anophthalmia/microphthalmia (Williamson & FitzPatrick 2014). Inhibiting the RA pathway with diethylaminobenzaldehyde (DEAB) in zebrafish induces a range of eye phenotypes (i.e. ranging from dorsalised retina to microphthalmia) depending on the time points it is administered and the dosage used (Le et al. 2012; Marsh-Armstrong et al. 1994). To date causative mutations have been identified in three genes that are components of the RA synthesis pathway; *STRA6, ALDH1A3* and *RARB* (Williamson & FitzPatrick 2014). The RA signalling cascade is unique among the developmental signalling pathways in that the signal is not carried by a protein but a simple organic molecule, retinoic acid. Retinoic acid is a derivative of vitamin A (retinol) which,if secreted,can act as a Ligand. Alternatively it can be activated within the cell that synthesised it. In this case RA is used for intracellular signalling (Duester 2008). Retinol is usually transported in the body bound to a carrier protein called RBP4 (Duester 2008).

STRA6 promotes the absorption of retinol-RBP4 (Kawaguchi et al. 2007). This increases the available supply of retinol for conversion into RA and thus acts as an agonist of RA signalling (Rhinn & Dollé 2012). RA signalling is known to be utilised throughout eye development **(Table 1.2.1)** and indeed in multiple developmental processes. This is shown by the fact that biallelic *STRA6* mutations cause pulmonary hyperplasia/diaphragmatic hernia/eventranation, anophthalmia /microphthalmia and cardiac defect (PDAC) syndrome (Chitayat et al. 2007). PDAC is a very distinctive syndromic cause of anophthalmia and is most often embryonic lethal (Chitayat et al. 2007). with 19 unrelated mutations reported (Williamson & FitzPatrick 2014).

Once the retinol has been absorbed it must be converted into RA. This reaction occurs in two stages. The retinol is first converted into retinaldehyde which is then converted into RA. ALDH1A3 is an enzyme that catalyses the conversion of retinaldehyde into RA (Hsu et al. 1994). Patients with homozygous *ALDH1A3* mutations display bilateral eye abnormalities, both anophthalmia and microphthalmia (Fares-Taie et al. 2013). To date 11 unrelated incidences of anophthalmia/microphthalmia families have been identified with biallelic *ALDH1A3* mutations(Williamson & FitzPatrick 2014).

Once retinoic acid is synthesised it facilitates the activation of target genes by binding target sequences. This process is accomplished by retinoic acid receptor (RAR) proteins. Once the retinoic acid binds to them, they are able to bind the target sites and up regulate target genes (Rhinn & Dollé 2012). RARB is a tissue specific member of the RAR family of proteins. Only four families have been identified with *RARB* mutations, which have been observed to be both monoallelic and biallelic (Srour et al. 2013).

There are several other plausible candidate genes which may account for a small percentage of anophthalmia/microphthalmia (Williamson & FitzPatrick 2014), However they remain less well studied, account for a small percentage of cases and/or are associated with less severe eye defects (Ranging from microphthalmia to coloboma; Williamson & FitzPatrick 2014). It is likely that several more will be identified in the future.

One recent example is *MAB21L2.* Mutations in *MAB21L2* only account for a tiny fraction of anophthalmia cases (five families have been identified; Rainger et al. 2014). What is noteworthy is that the majority of them seem to be residue specific (Rainger et al. 2014). Four out of the five mutations identified are monoallelic and have occurred independently within 2 residues of each other (three in Arg51 and one in Glu49). The one remaining mutation occurs in Arg247, and despite being biallelic the phenotype is substantially less severe. The siblings have bilateral coloboma (Rainger et al. 2014). Little is known about the function of MAB21L2; the wild type protein has been shown to bind to RNA *in vitro* and this binding is lost in the mutant alleles (Rainger et al. 2014). It is reasonable to speculate that it is this loss of binding of target RNAs that leads to anophthalmia but further work is needed to confirm this.

While none of the genes discussed have been shown to be causative genes for OAS they all represent plausible potential interactors with SMOC-1.

## 1.2.2 Potential causes of anophthalmia

There are broadly three classes of anophthalmia. Primary anophthalmia, where the eye fails to develop at all. This is due to the failure of the establishment of the eye field andoccurs extremely early in development (Fitzpatrick & van Heyningen 2005). If SMOC1 controls the shape of signalling gradients (e.g. BMP, SHH or RA) then it could affect the size of the domains and even presence of early eyefield master regulators (e.g. PAX6, SOX2 and OTX2) which would lead to an absence of any other eye tissues, as is the case in *Xenopus rx* morphants (Mathers et al. 1997).

The second possibility would be regressive anophthalmia, where the eye field begins to develop though at some point eye development fails. In this scenario, an arrest in eye development may result in eye degeneration, casuing true anophthalmia. This is the most common mechanism of eyelessness in blind, cave dwelling vertebrate (fish?), though the molecular details vary between species. For example the eyes of blind cave *Astyanaxs* undergo broadly normal eye development up to 24 hpf , only for the lens to become apoptotic (Hinaux et al. 2011). Without the signals from the lens the RPE begins to degrade resulting in anophthalmic fish. In this example the degradation is lens dependent. If a lens is transplanted into the embryo from a sighted species of *Astyanaxs* into a blind *Astyanaxs,* then the fish will go on to develop healthy eyes and when you transplant the lens of a blind *Astyanaxs* embryo into a sighted *Astyanaxs* that embryo goes on to develop anophthalmia (Yamamoto & Jeffery 2000). Whilst the

precise molecular mechanism is unknown, it is thought to be mediated by an upregulation of hedgehog signalling during gastrulation (Yamamoto et al. 2004).

The final possibility is Secondary anophthalmia where eye development arrests at an early stage of development resulting in small rudimentary eyes that only become apparent upon detailed analysis. This has also been observed in blind cave-dwelling animals. For example in *Poecilia mexicana* eye development is arrested immediately after eye formation resulting a significant reduction in eye size (Yamamoto et al. 2004).

Of the three competing explanations of the anophthalmia phenotype of OAS patients we can safely discount secondary anophthalmia as OAS patients often display true anophthalmia and secondary anophthalmia would leave residual optic tissue(Fitzpatrick & van Heyningen 2005). Of the remaining options regression seems more likely as it is far more common in nature. Every known blind cave vertebrate undergoes some level of eye development even though the eyes are nonfunctional. This would seem like a very evolutionarily inefficient system. Why would it be beneficial for an organism to spend the energy forming eyes that are never used? The answer is thought to be that in vertebrates the formation of the eyefield is inextricably linked to the formation of the telencephalon (Woo & Fraser 1995; Woo et al. 1995) and any primary anophthalmia in a vertebrate would also be associated with severe defects in brain development (Rétaux & Casane 2013). The counter argument to this is that OAS is associated with learning difficulties, but the penetrance is far lower than you would expect if the anophthalmia was primarily due to primary anophthalmia.

# 1.3. Main Aims

While it has become clear that mutations in *SMOC1* are the principal cause of *OAS* the molecular function of *SMOC1* is still poorly understood. The overall aim of my thesis is to help elucidate the function of *SMOC1*, and through doing this understand how disrupting these processes leads to the clinical features of OAS. To do this I hoped to address two central questions.

### 1.3.1 How do defects in *SMOC1* function cause the features of OAS?

To answer this question I aim to establish a zebrafish model of OAS. Principally because the ex-vivo development of zebrafish allowed me to observe the effects of SMOC1 depletion on the formation of the eye in real-time.

### 1.3.2 Establish if smoc-1 is functionally homologous to *Drosophila* pentagone

At the outset of the thesis, the best models for SMOC1 function had implicated it as a simple antagonist of BMP signalling. However just prior to the beginning of the project, work on the *Drosophila* orthologue of SMOC1 (pentagone) proposed a more complicated model. Pentagone was proposed to act by controlling the shape of (dpp/BMP) signalling gradient, and whether or not it acts to antagonise or agonise the (dpp/BMP) signalling pathway depends on the context of the perturbation. This model was further validated by the work of Hamaratoglu et al. 2011.

The second aim of my thesis is to confirm whether or not this model is conserved in vertebrates, and more specifically in zebrafish.

# Chapter 2

Chapter 2 will focus on establishing that zebrafish are a valid model organism for studying *SMOC1*. I will first demonstrate that zebrafish do include an intact copy of the *zsmoc1* gene, and that this gene is expressed during zebrafish development, i.e. within the first 24 hours of zebrafish development. I will confirm the presence of any alternative splicing, both to identify any biologically relevant alternative transcripts but also aid in the targeting of endonucleases to produce true knockout zebrafish lines.

Once I identified DNA sequences it will be analysed so we can predict functional significance of different zsmoc-1 domains. While this information is of biological significance it can also be used to target endonucleases more effectively. In summary I aim to:

### 2.1 Define the correct gene structure of *zsmoc1*

I aimed to determine whether the apparently rearranged gene structure of *zsmoc1* in the current version of the genome assembly was correct.

### 2.2 Establish if any alternative splicing is present in *zsmoc1*.

Alternative splicing is widespread, with many eukaryotic genes being alternatively spliced. I aimed to determine the splice variants in *zsmoc1* to enable accurate

targeting of TALENs and morpholino reagents for efficient depletion of *zsmoc1* throughout development.

**2. 3 Analyse the structure of the zsmoc-1 protein product bioinformatically.**

**2.4 Use cross species analysis to search for novel human *SMOC1* transcript variants.**

## Chapter 3

Chapter 3 describes the use of morpholino oligonucleotides targeting *zsmoc1* to:

**3. 1 Determine if *zsmoc1* morphants are truly dorsalised.**

Pilot work undertaken in my masters project showed that zebrafish appeared to display a characteristic dorsalised phenotype of a bmp antagonist which implies an early loss of BMP signalling. If *zsmoc1* morphants are dorsalised they should show an increase in the levels of dorsal marker genes and a corresponding decrease in the levels of ventral marker genes.

To establish this, a novel qPCR quantitative assay of embryo dorsalisation /ventralisation will be developed. The assay will be developed using qPCR to establish the relative expression level of marker genes between 12 and 16 hpf. Once validated using morpholinos which directly alter bmp signalling, the assay will be used to assess embryos in which *zsmoc1* has been knocked-down.

### 3. 2 Characterise *zsmoc1* morpholinos

Morpholino oligonucleotides are known to have quite severe off target effects. To rule this out, I will take two approaches: First, inhibit the p53 cell death pathway with co-injection of *zsmoc1* targeting MO in any morpholino work as recommended in Huang et al. (2012) as this is a pathway most often connected with morpholino toxicity (Bedell et al. 2011). Second, *hSMOC1* will be cloned into a CS2+ plasmid backbone to enable the production of capped mRNA which can be used to try to rescue the *zsmoc1* morphant phenotype.

### 3. 3 Use real-time imaging of fluorescent transgenic reporter fish to understand the developmental pathology induced by *zsmoc1* morpholinos.

To achieve this an automated macroscope system will be optimised so both the establishment of the eyefield and the development of the eye can be monitored in real time from between 12 and 24 hpf.

Given the severe eye defects associated with OAS and the fact that *zsmoc1* is expressed in the developing zebrafish eye, and also as bmp signalling is involved in the establishment of the eyefield, it is plausible that *zsmoc1* is involved in regulating this gradient. If this is the case then *zsmoc1* morphants should display eye defects, which should manifest is a reduction in eyefield size or a delay in eyefield development. To accurately assess zebrafish eyefield size, real-time

imaging of the rx3:gfp (Rembold et al. 2006) zebrafish line will be used as *rx3* is the earliest known marker of the zebrafish eye field.

Additionally there is a bmp signaling gradient which is established within the eye field at between 16 hpf that establishes the dorsal ventral pattern across the optic cup. The signalling gradient is confined from dorsal to ventral. Given that *zsmoc1* is expressed within ventral optic cup it is a plausible regulator of the dorso-ventral bmp gradient.

If this is the case then the dorsoventral gradient should be disrupted. To test this a BRE:gfp (Collery & Link 2011) bmp reporter line will be used to visualise the area of bmp activity on both *zsmoc1* morphant and control embryos.

### 3. 4 Establish if *zsmoc1* acts as a *Pentagone* homologue

Whole mount immunostaining and OPT quantitative image analysis will be used to determine if p-smad gradient is established in the same manner as the mad gradient in *Drosophila* (Vuilleumier, R. et al. 2010). Overexpression of *smoc1* mRNA should extend the range of p-smad activation but weaken the signal, while injection of zsmoc1MO should result in a contracted gradient but have stronger p-smad staining.

# Chapter 4

While the morpholino work is compelling, **(Chapter 3)** the repetition of the results with a genetic inactivation of *zsmoc1* is highly desirable as it would conclusively rule out off target effects.

In summary I aim to:

### 4.1 Induce loss of function genomic mutations in *zsmoc1*

To this end I decided to follow a TALEN based strategy.

Which required:

#### 4.1. 1 Design multiple TALEN pairs that target *zsmoc1*.

#### 4.1. 2 Constructing the TALEN construct using Golden Gate cloning.

#### 4.1. 3 Confirm that the synthesis was successful and the constructs cleave DNA *in vitro*.

#### 4.1.4 Establish whether TALEN pairs induce hereditable mutations *in vivo*.

# Chapter 2

# Characterising the *smoc1* locus in zebrafish

# 2.1 Introduction

## 2.1.1 *Danio rerio*

*Danio rerio,* more commonly known as zebrafish, have been studied for over 70 years. Following the work of Streisinger et al. in the late 70s and early 80s zebrafish have become an extremely popular developmental organism. Used alongside mice, chicks, *Xenopus* and *Drosophila*.

Zebrafish are carnivorous cyprinids, native to southern India. They probably originally evolved to survive on floodplains in shallow pools of water, but the prevalence of irrigation used in the cultivation of rice by humans has proved an ideal habitat for zebrafish, and where they are endemic. It is a small hardy fish, no more than 4 cm long with characteristic blue horizontal stripes from which they get their name (Spence et al. 2008; Engeszer et al. 2007).

## 2.1.2 *Danio rerio* as a useful model of disease

Zebrafish have a number of characteristics which make them an attractive model organism particularly for developmental biology. Firstly they are small and robust fish that can be kept cheaply in large numbers. While their gestation time is similar to mice, between 3-4 months they are prolific breeders with a single mating pair producing between 100 and 300 embryos every 7-14 days (Spence et al. 2008). This enables high throughput experiments and large screens (e.g. induced mutagenesis or small molecule/chemical biology). The embryos develop rapidly; zebrafish embryos start feeding after only five days and most of the organogenesis is completed after only 36 hours (Kimmel et al. 1995).

They develop ex vivo and the embryos are translucent, making them ideal for imaging embryonic development. These characteristics allow fluorescent reporter transgenic animals (e.g. GFP tagged proteins) to be used to visualise zebrafish gene expression in real-time (Gosse & Baier 2009; Collery & Link 2011; Molina et al. 2007) and resolve individual organs or even single cells **(Fig. 3.2.14;** Asai-Coakwell et al. 2009).

While zebrafish are not as closely related to humans as other model organisms (e.g. mice), they do have orthologues to the majority of human genes (71.2%). However only half of these genes are direct one-to-one homologues due to genome duplication (Howe et al. 2013). Whilst care should be taken when extrapolating results from zebrafish to human diseases they have already provided insights into several human diseases. Driever et al. and Haffter et al. conducted the

first large scale screens of zebrafish and identified several mutations that proved to be analogous to human disease (Driever et al. 1996; Haffter et al. 1996). For example the zebrafish mutant *sauternes* was mapped to ALAS2. Once the phenotype was characterised it went on to be the first animal model for congenital sideroblastic anaemia (Brownlie et al. 1998). The mutant *gridlock* has been used as a model for aortic coarctation, which has led to the identification of compounds that suppress the developmental pathology leading to coarctation (Peterson et al. 2004).

Even in cases where the precise phenotype is not conserved, model organisms can provide insights into human biology. Biological networks are more strongly conserved; interactions identified in one model species strongly predict interaction in another species, even if the phenotype is different (McGary et al. 2010). Bmp signalling is a classic example of this. In zebrafish defects in the bmp signalling pathway during gastrulation trigger a distinctive spectrum of phenotypes characterised in Kishimoto et al. (1997; **Fig. 3.1.3)**. These phenotypes were induced by varying levels of bmp activity. Using them it was possible to rapidly screen zebrafish embryos for mutation in bmp interactors. The vast majority of interactors have conserved roles up to mammals despite having different phenotypes. For example R-smads (Transcription factor; Fujii et al. 1999; Dick et al. 1999), BAMBI (repressor) (Tsang et al. 2000;Grotewold et al. 2001) and Chordin(repressor) (Branam et al. 2010) have all been shown to have conserved functions in multiple model organisms (reviewed in Ramel & Hill 2012). Of course several novel proteins unique to zebrafish were identified as well but they

appeared to play a relatively minor role e.g. the zebrafish homeobox gene Dharma is an antagonist of bmp2b (Koos & Ho 1999).

Given their advantages in imaging, zebrafish embryos are used extensively in developmental biology. For example Zebrafish mutational screens identified numerous phenotypes (Driever et al. 1996; Da Costa et al. 2014; Moens et al. 2008; ; Wang et al. 2007). A good example of this was the use of this system to identify *one eyed pinhead*, a member of the nodal pathway and a significant player in early development (Gritsman et al. 1999).

Zebrafish are increasingly used to validate that mutations identified in humans are causative. For example if a missense variant fails to rescue a developmental effect induced by loss of function of the orthologous gene in fish embryos when wildtype mRNA can it is considered a null (Manzini et al. 2012; Sivasubbu et al. 2013). This approach has even been used in large scale assays to quantify the relative functional activity of multiple variants of the same gene by comparing the degree to which they rescue the Morpholino knockdown of their zebrafish orthologue (Kabashi et al. 2011). Zebrafish are also valuable model organisms for drug discovery. For example Yeh et al. conducted a small molecule screen which identified several novel inhibitors of angiogenesis, which have potential as anti-cancer drugs. They went on to implicate two previously ignored pathways (both the COX-2-dependent signalling and beta-catinin signalling ) as targets for the development of other drugs (Yeh et al. 2009).

## 2.1.3 Zebrafish Genomics

The zygotes of zebrafish tolerate microinjection and the levels of gene products within the early embryo can thus be easily modified via the use of morpholino oligonucleotides (MO). MO are short modified oligonucleotides resistant to degradation. In zebrafish studies they are commonly designed to be complementary to mRNA. If targeted correctly they can impede either mRNA translation or splicing (Nasevicius & Ekker 2000; Morcos 2007; Bedell et al. 2011). Over-expression of a desired gene can be achieved by injecting a plasmid with zebrafish promoters but since DNA transcription only begins at the blastula stage of zebrafish (4hpf; Skromne & Prince 2008) this generally produces mosaic expression patterns (Kabashi et al. 2011; Monte Westerfield 2000). A more common strategy is to inject artificially synthesized mRNA directly into a single cell embryo which often leads to more even and ubiquitous expression (Finckbeiner et al. 2011; Skromne & Prince 2008).

Gene targeting/knockout technology in zebrafish has lagged behind mice as zebrafish appear to lack a homologous recombination pathway. Thus targeted mutations must be directly induced at the target site via technologies such as zinc finger nucleases (ZFNs; Ekker 2008), transcription activator-like effector nucleases (TALENs; Cade et al. 2012) and clustered regularly interspaced short palindromic repeats (CRISPRs; Hruscha et al. 2013). However, a wide array of mutant lines available in specific genes have been identified via TILLING (targeting induced local lesions in genomes) projects. Zebrafish are subjected to large-scale random

mutagenesis (typically chemically induced by ENU) and high throughput sequencing is used to identify mutations in individual genes (Stemple 2004; Moens et al. 2008). For example Tilling was used to generate mutations in *superoxide dismutase gene* (*sod1*) which will then be used to generate a better model of motor neuron disease (MND) which more closely mimics the features of MND (Da Costa et al. 2014).

The other obstacle is the current state of the zebrafish genome. The sequencing project was started in 2001 and was finished in 2013 (Howe et al. 2013). In the most recent genome build 83% is comprised of sequence mapped by clone by clone sequencing(Howe et al. 2013). This is where 50-60 kb sections of a genome are cloned into bacterial artificial chromosomes (BACs) before the whole segment is sequenced. Once the full 50-60 kb scaffold is sequenced it is then assembled with the other scaffolds (Brown TA. 2002). The remaining 17% of the zebrafish genome was assembled by more error prone shotgun sequencing where the whole genome is fragmented and sequenced directly(Howe et al. 2013; Staden 1979; Anderson 1981).

Adequate alignment and annotation of the zebrafish genome proved challenging as it is a member of the Teleostei infraclass of fish. The common ancestor of this class went through the Teleost-specific genome duplication (TSD), which comprised a whole genome duplication event specific to Teleost fish (Meyer & Schartl 1999;Taylor et al. 2003).

Genome duplication presents a problem for genomic sequencing because it means the zebrafish genome is highly repetitive. As sequencing reads can only be so long (~900bp) they must be aligned together. Any repetition that is both

sufficiently long and similar can be misaligned. Either multiple distinct paralogous sequences can be erroneously combined into a single loci, or paralogues can be missed because they are mapped to the wrong loci **(Fig. 2.1.1;** Eichler 2001).

If the repetitive sequence is greater than 1 kb both clone by clone sequencing and shotgun sequencing can make these kinds of mistakes. Though shotgun sequencing is far more error prone because the whole genome is sequenced in one step which means any repetitive elements in the entire genome can lead to assembly errors. While with clone by clone sequencing the repetitive elements must either be located within the same scaffold (50-60 kb) or at the beginning or end of a scaffold(Brown TA. 2002).

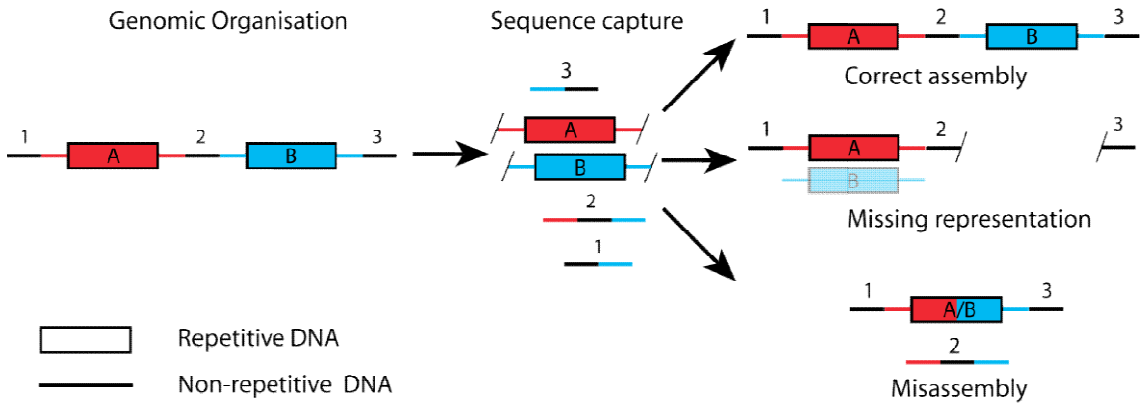



**Fig 2.1.1 Schematic representation of different outcomes of genomic assembly.**

If a sequence of repetitive DNA is not correctly aligned it is possible for software to misalign the sequence in 2 ways. 1. Misrepresentation, or 2. Misassembly.

Figure based on (Eichler 2001).

The zebrafish genome is the most highly repetitive animal genome sequenced so far which explains why it has taken 9 years longer to finish the zebrafish genome than originally predicted. For comparison the *Xenopus tropicalis* started in 2002 and was finished by 2011 despite having a larger genome **(Table 2.1.1;** Hellsten et al. 2010**).** Even though the genome has been officially finished over 4% of the scaffolds have not been mapped to the reference genome. Though the Genome Reference Consortium continues to update the zebrafish genome and one day these maps will be resolved.

| Species | start date | end date | Size | Protein coding genes | Completion |
|---|---|---|---|---|---|
| *Danio Rerio v9.4* | 2001 | 2013 | 1.4 x 109 | 26,000 | 94.00% |
| *Xenopus tropicalis v.4.1* | 2002 | 2010 | 1.7 x 109 | 21,000 | 97.60% |

**Table2.1.1 Summary of the zebrafish and *Xenopus* genome projects. (**data taken from Howe et al. 2013; Hellsten et al. 2010)

# 2.2 Results

## 2.2.1 Clarifying the Genomic organization of *zsmoc1*

To further the understanding of *smoc1* zebrafish was chosen as a model organism for the reasons outlined above. The first challenge involved identifying the zebrafish orthologue of *smoc1*. Using the human *SMOC1* cDNA two zebrafish sequences were identified by BLAT searching of UCSC genome browser Jul 2010 Zv9 dan Rer 7. However, the genomic sequence annotates *zsmoc1* as two adjacent fragments arranged in opposing orientations on chromosome 17 **(Fig. 2.2.1. A).** One fragment is located between genomic coordinates 48424231 and 48449106. This is transcribed in the reverse orientation and contains exons 2 to 9. The other fragment is located between coordinates 48482432 and 48542881. It is comprised of exons 10 to 15, and is transcribed in a forward orientation. Together these genes structures account for the full-length of vertebrate *smoc1*.

Yet, if this represents the true genomic organisation of *zsmoc1* then these are probably pseudogenes of no functional relevance. To establish whether the *zsmoc1* locus in the current version of the zebrafish genome is correct, RT-PCR analysis was performed. PCR primers were designed flanking exons 8 and 9 junction and zebrafish cDNA was used as a template **(Fig. 2.2.1. A,B).** This demonstrated that it was possible to PCR across the exon 8/9 junction, thus

proving that full-length *zsmoc1* transcript is present in vivo. The *zsmoc1* locus in this genome assembly has been misassembled.

A further BLAT search identified a zebrafish contig NW_001879596.1 based on Zv7 _NA65. This sequence clearly encompasses the intron between exons nine and ten, and shows that the intron is approximately 11,000 bp long. This data leads to the new genome map shown in **Fig. 2.2.1 C**. However, it is yet to be confirmed experimentally by long range PCR.

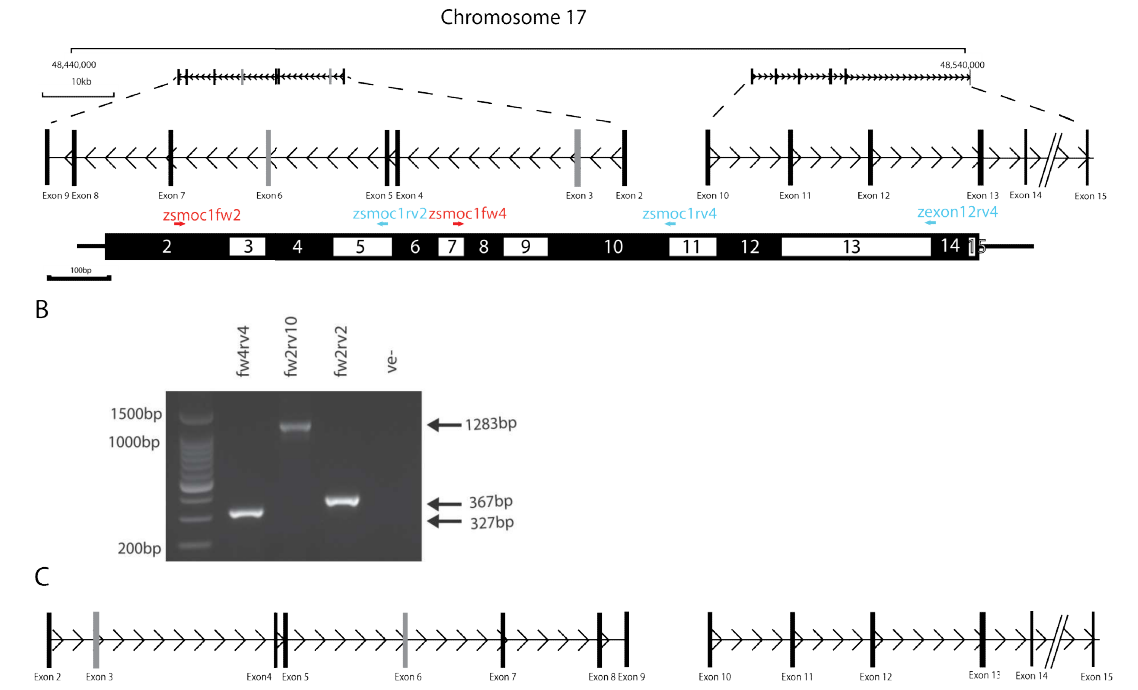



**Fig. 2.2.1 The genomic organisation of the zebrafish *smoc1* gene.**

***(A)*** Schematic representation of the Jul 2010 Zv9 dan Rer 7 build of the zebrafish genome on UCSC genome browser showing that zsmoc1 is recorded in two separate genes, which are adjacent but transcribed in opposing orientations. Grey exons represent exons only found in zebrafish. The structure of expressed *zsmoc1*, exons are numbers and labeled alternately black and white. UTRs and shown by black lines. The primers used in RT analysis are labeled with arrows, blue for forward and red for reverse. **(B)** RTPCR analysis of the exon 8/9 junctions form 24 hpf zebrafish cDNA showing that fragments can amplified across the 8/9 junction. **(C)** A corrected genomic locus structure of the smoc1 based on NW_001879596.1 Zv7 _NA65.

## 2.2.2 The recorded sequence of zebrafish *zsmoc-1* is incomplete

Once the genomic structure of *zsmoc1* was corrected I was able to amplify and clone the apparent full-length cDNA. This was then used to synthesise *zsmoc1* mRNA. Injection of *zsmoc1* mRNA into zebrafish zygotes resulted in no overexpression phenotype (data not shown). This was surprising because in *Xenopus xsmoc1* overexpression results in severe ventralisation (Thomas et al. 2009) and raised the possibility the current *zsmoc1* cDNA was missing the signalling peptide sequence, as this did not appear to be conserved between zebrafish and mammals.

Signalling peptides are short stretches of 5-30 aa that have a high hydrophobicity. They induce cleavage by signal peptidase which places the protein into the secretory pathway (von Heijne & Gavel 1988). To investigate this, the signalling peptides of an array of SMOC-1 proteins from several species were compared using SignalP v4.1 **(Fig. 2.2.2).** The first 50 amino acids of SMOC-1 were compared between zebrafish, human, mouse, *and Xenopus*. *Drosophila* Pentagone, the fly homologue of the SMOC proteins was also included in the analysis. SignalP v4.1 (Petersen et al. 2011) was used to identify a 20 amino acids region of high hydrophobicity in every Smoc-1 protein except for in zebrafish. It also identified the target sequence for signal peptidase at the end of this hydrophobic region for each protein excluding zsmoc-1. This suggested that either zsmoc-1 is not secreted or that the zebrafish cDNA sequence is missing the signal peptide sequence due to misassignment of the initiating methionine.

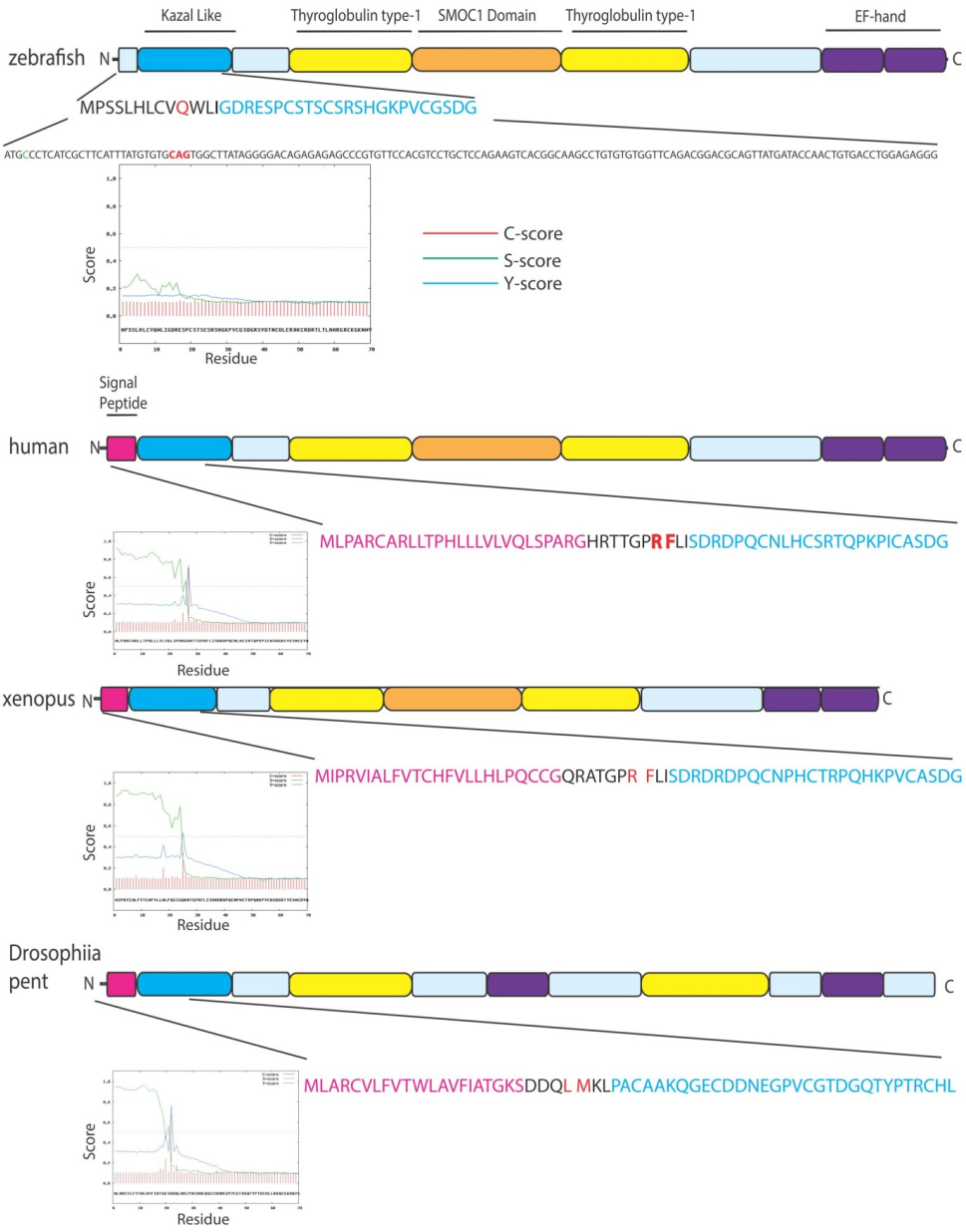



**Fig. 2.2.2 A comparison of signal peptide prediction algorithms on SMOC-1 in multiple species.**

The first 60 amino acids predicted in smoc-1 in zebrafish, humans and *Xenopus* were analysed using signalIP v4.1 along with the *Drosophila* homologue of smoc1 pent. Each amino acid was given a C-score (red) S-Score(green) and Y-score(blue). A high C-score denotes the first amino acid sequence after cleavage occurs. A high S-Score shows the sequence with high signal peptide identity. Y-score is the combined scores S and C scores. All proteins analysed with the exception of zsmoc-1 showed high cleavage score(they have Matthews correlations of greater than 0.5 for C-score or Y-score) 20 residues downstream of their start codons. The first ~20 residues of every protein analysed have high signal peptide identity (a S-Scores<0.5) with the exception of zsmoc-1. Taken together only zsmoc-1 lacks a signal peptide prediction. Each protein shows a coding region of the first 60 aa. The predicted signalling peptide is depicted in pink, the Kazal domain is depicted in the blue and splice sites are depicted in red. In the human , *Xenopus* and *Drosophila* proteins analysed the signalling peptide was spliced onto the mature transcript three residues upstream of the Kazel domain(highlighted in red). The predicted zebrafish protein has the splice acceptor sequence 9bp(3 aa) upstream of the Kazel domain(highlighted in red).

To further investigate this, the exon structure of the proteins was compared across species. In all Smoc-1 proteins and in *Drosophila* Pentagone the signal peptide was alternatively spliced onto the protein three amino acids before the Kazal domain. Analysis of the zebrafish sequence, revealed a canonical splice acceptor sequence (CAG) 3 amino acids before the Kazal domain, strongly indicating that there is another SMOC1 variant present in zebrafish where the signalling peptide is spliced onto zsmoc-1.

## 2.2.3 5'RACE identifies the signal peptide open reading frame and 5’ UTR of *zsmoc1*

Alignment of the genomic structures of *SMOC1* in vertebrates revealed that the first intron is always large, ranging from 50,000 to 80,000bp **(Sup. Table 2.)**. This makes a computational search for the start codon impractical. Instead the five prime UTR of *zsmoc1* was identified by 5' RLM RACE. Total RNA was isolated from 12hpf AB embryos before being processed by 5' RLM RACE. Due to the low efficiency of RACE PCR and the relatively low abundance of *zsmoc1* in zebrafish, the transcript was amplified sequentially with two nested PCRs. An initial outer PCR produced no bands, but this is not uncommon for RACE PCRs, as only the most abundant transcripts produce bands in the initial stages. The second inner PCR produced three bands. The band in each of the two *zsmoc1* lanes were both the same size (~450bp) indicating that there is likely to be a single 5'UTR in *zsmoc1* at

12hpf. There was one unexpected band of a different size in the positive control *bmp2b* secondary amplification. This band was sequenced and shown to be a spurious amplification of the seventh exon of *rasa3*(data not shown).

To confirm the identities of the 450bp band, this was excised from both lanes on the gel, TA cloned and sequenced **(Fig. 2.2.3 C).** BLAT Searching this sequence against the zebrafish genome (NW_001879596.1 Zv7 _NA65) showed that the entire fragment maps close to *zsmoc1* on chromosome 17**(Sup. Fig. 3)** between positions 48,480,000 and 48,600,000. 39 bp maps to the current first exon of the *zsmoc1* gene while the remaining sequence maps to an unannotated region. To test whether this fragment coded for the missing signal peptide the sequence was translated with the ExPASy translate tool. A 33 residue open reading frame was discovered in the same frame as *zsmoc1*. SignalIP V4 predicts that this open reading frame encodes a signal peptide **(Fig. 2.2.3 D)** although the Y score is just below the threshold for significance(which was taken to be a Matthew Pearson correlation of lower than 0.5; **chapter 6 6.5.6**). However the results are still compelling as the predicted cleavage site is located at the exact residues where it is found in other vertebrate smoc-1s **(Fig. 2.2.2 )**.

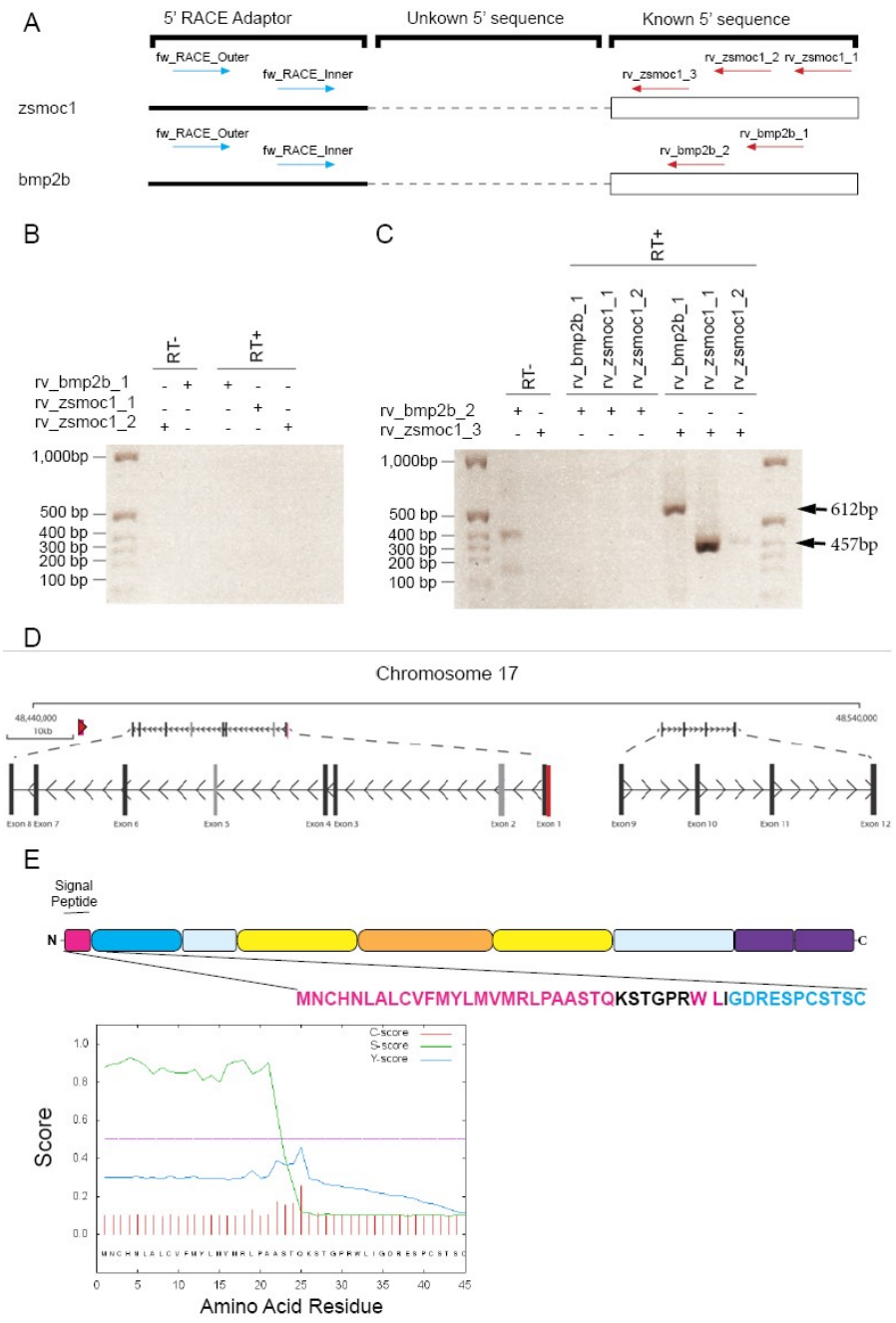



**Fig. 2.2.3 5'RACE identified a novel 5' UTR exon coding the signal peptide of *zsmoc1.***

**(A)** Diagrammatic representation of the 2 step nested RACE PCR used to isolate 5' region of *zsmoc1*. The primers used are labeled with arrows, blue for forward and red for reverse **(B)**The outer RLM-RACE-PCR amplification of both zsmoc1 targeting primers and bmp2b positive control primers were free of detectable bands. **(C)** Analysis of The inner RLM-RACE-PCR did generate detectable bands in both of the predicted lanes at 457bp(marked by arrow) but also a second larger band was generated in the negative control for the rv_bmp2b_sp1 (612bp marked by arrow). All bands were excised, cloned and sequenced. The sequences generated were compared to the reference zebrafish genome V5 using BLAT. The sequence of the 457bp fragment proved to match precisely to regions of the zebrafish genome. One located at the beginning of the first exon of the known *zsmoc1* gene. The other located 10kb downstream of exon 8 in the reverse orientation **(D).** The unexplained band at 612bp proved to be an amplification of an unrelated zebrafish scaffold protein which shared some homology with the bmp2b primers. When the *zsmoc1* sequence was translated and the open reading frame analysed using signalIP V4.0 it clearly identified the signalling peptide **(E).**

## 2.2.4 Identifying transcript variants of *zsmoc1*

The RT-PCR using the primers spanning exons 6 to 8 consistently returned multiple bands **(Fig. 2.2.4 C).** This raises the possibility that *zsmoc1* has multiple isoforms. To test this these PCR product were cloned into pGEM®-T Easy Vector using TA cloning. Ninety-six clones were sequenced. From these sequences it became apparent that there were seven different *zsmoc1* sequences **(Fig. 2.2.4) (Fig. 2.2.5).** Six of the sequences are previously unrecognised *zsmoc1* transcripts. All alternative splicing was confined to the SMOC domain. The SMOC domain is a unique functional domain identified in the original paper that identified SMOC-1 (Vannahme et al. 2002). It is located between the thyroglobulin domains. The region has no known function but is conserved in all vertebrates in both SMOC-1 and SMOC-2. The number of transcripts are quantified and the relative abundance is displayed as percentages **(Fig. 2.2.4 B; Fig. 2.2.5 B)**.

Alternative isoforms in different genes have a range of known functions including changing the stability, localisation and catalytic activity of protein (Kelemen et al. 2013; Stamm et al. 2005). In other cases any functional significance of an alternative splice form is unknown. This is particularly the case for low abundance transcripts in which aberrant splicing may be caused by stochastic errors within the splicing machinery(Kelemen et al. 2013; Stamm et al. 2005). To decrease the chances of pursuing on non-functional or functionally redundant variants, I focused further analysis on the highly abundant *zsmoc1* transcripts. High abundance transcripts were defined here as variants that made up over 10%

of the total population of clones at 24hpf and are detailed in **Fig. 2.2.4**. While the remaining isoforms were classified as minor isoforms and are detailed in **Fig. 2.2.5.**

The isoforms are labeled A to G in order of relative abundance. Isoform A was the most abundant transcript in the clonal analysis, so was taken to be the canonical *zsmoc1* sequence. Each of the identified splice forms maintain the open reading frame of *zsmoc1*. All the identified splice sites also follow the "GU-AG" rule of splicing **(Fig. 2.2.4. A; Fig. 2.3.5. A).** These features support the hypothesis that these transcripts are not the spurious products of aberrant splicing (Magen & Ast 2005; Xing & Lee 2005).

Isoform B omits exon 6, which resulted in the omission of 51 base pairs (corresponding to 17 amino acids) from the canonical sequence while Isoform C has an alternative splice acceptor site on exon 8 (8a), which results in the inclusion of an extra 33 bp (11 amino acid residues).

The minor *smoc1* splice variants are also each located within the SMOC domain. Variants E and F both involve an alternative splice donor site within exon 8(8'') resulting in a loss of 6bp (2aa). Variant E also skips exon six like Variant B. Variant F uses the same alternative splice donor site within intron 8 as used in Variant C. Variant D and Variant G are both the product of exon exclusion. In Variant D the transcript lacks exons 6 through 10 resulting in the complete absence of the smoc1 domain. Variant G lacks exons 7 and 8.

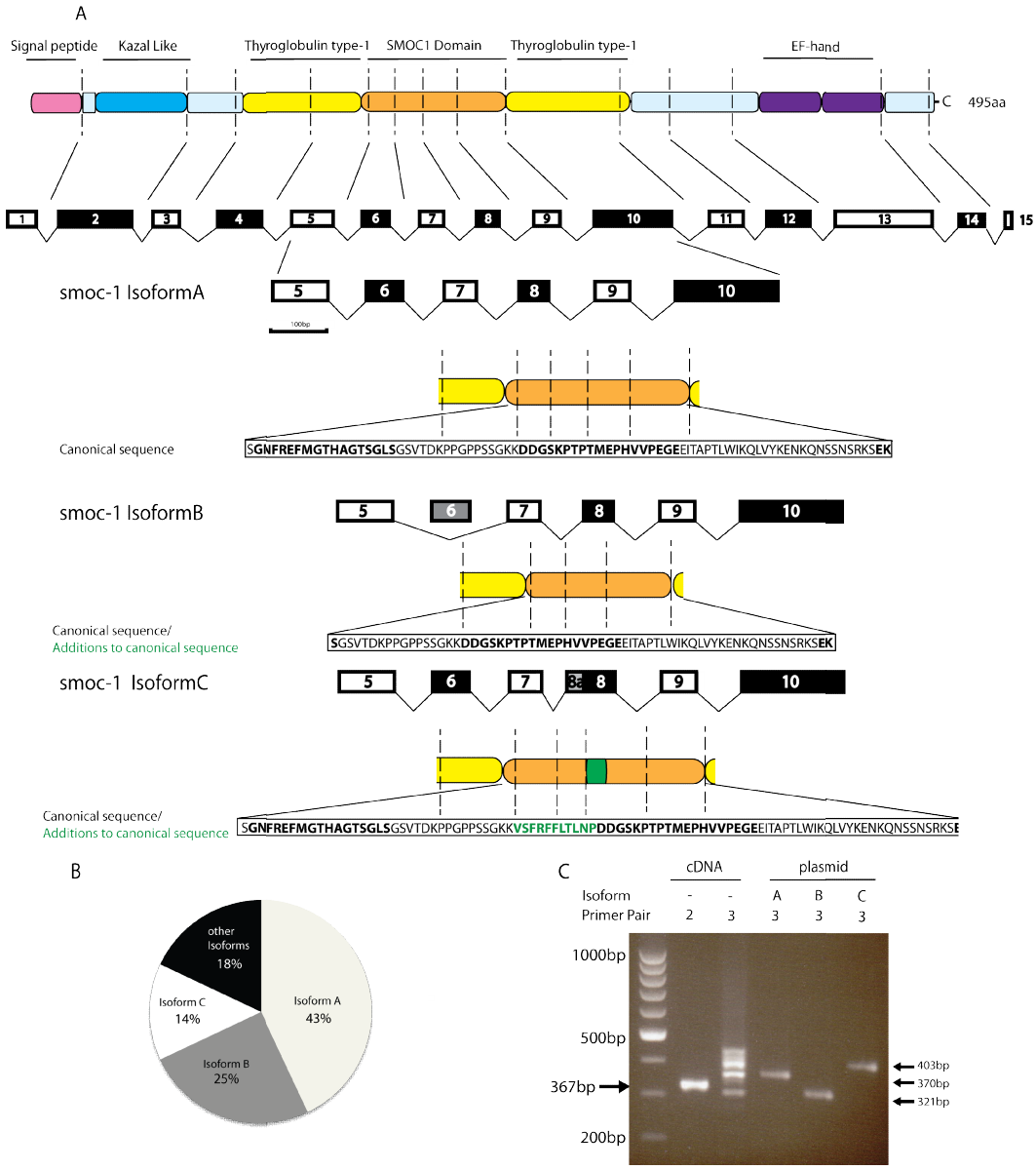



**Fig. 2.2.4 The structure of the zebrafish *smoc1* gene and its most common isoforms.**

**(A)** A schematic representation of the consensus transcript and the regions of the protein that each exon codes for. The consensus transcript is 495 amino acids long and coded for by 15 exons (black and white numbered boxes). All known functional domains are annotated. The exon structure of the three most common variants are shown. All identified alternative splicing was confined to the SMOC domain. Isoform B skips the splice acceptor site exon 6 which results in the omission of 51bp (or 17 aa) within the SMOC domain from the transcript. Isoform C has an alternative splice acceptor site upstream of exon seven resulting in the inclusion of an additional 33bp (11aa) pairs between exons six and 7 within the SMOC domain which has been highlighted in green. **(B)** The relative abundance of the different smoc1 isoforms measured by clonal analysis. The most abundant transcript was this isoform A (42%) so this was labeled the canonical zebrafish transcript. The next most abundant transcript was isoform B (26%) followed by isoform C (14%). The remaining transcripts were for isoforms which made up less than 10% of the population (18%). **(C)** RT-PCR analysis shows that amplification across the SMOC domain in cDNA results in multiple bands that can be attributed to alternative splicing. The three brightest bands directly correspond to the bands amplified from clones of isoforms A, B and C.

To confirm the presence of all the major transcripts RT-PCR analysis was performed on RNA from wild type zebrafish embryos. The three bands produced were of the predicted sizes for Variants A, B and C, and were the brightest bands of RT-PCR conducted using cDNA. This confirms that the isoforms both exist and represent the most abundant isoforms of *zsmoc1*. **(Fig. 2.2.4. C)** Other bands are observed which may represent the minor alternative splice forms detailed in **Fig. 2.2.5.** However, some of these bands probably represent heteroduplex formation given the mixed population of amplicons (Thompson et al. 2002). It is particularly prevalent between the variants that differ only slightly e.g. splice variants A, B and C, which differ by regions of either 33 bp or 51 bp. This difference in size leads to the formation of secondary structure within heteroduplexes (either a loop in the DNA or failing that a Y-shaped DNA structure) causing reduced mobility. The presence of heteroduplexes would explain why there are bands with an apparent size greater than splice variant C, the largest splice variant identified in our clonal analysis.

To test this, the PCR was repeated in the products were run out on polyacrylamide gels. As a reducing gel any DNA heteroduplexes would be denatured. The unexplained bands disappeared, proving they were an artifact (data not shown).

To summarise seven *zsmoc1* isoforms have been identified, of which six are novel.

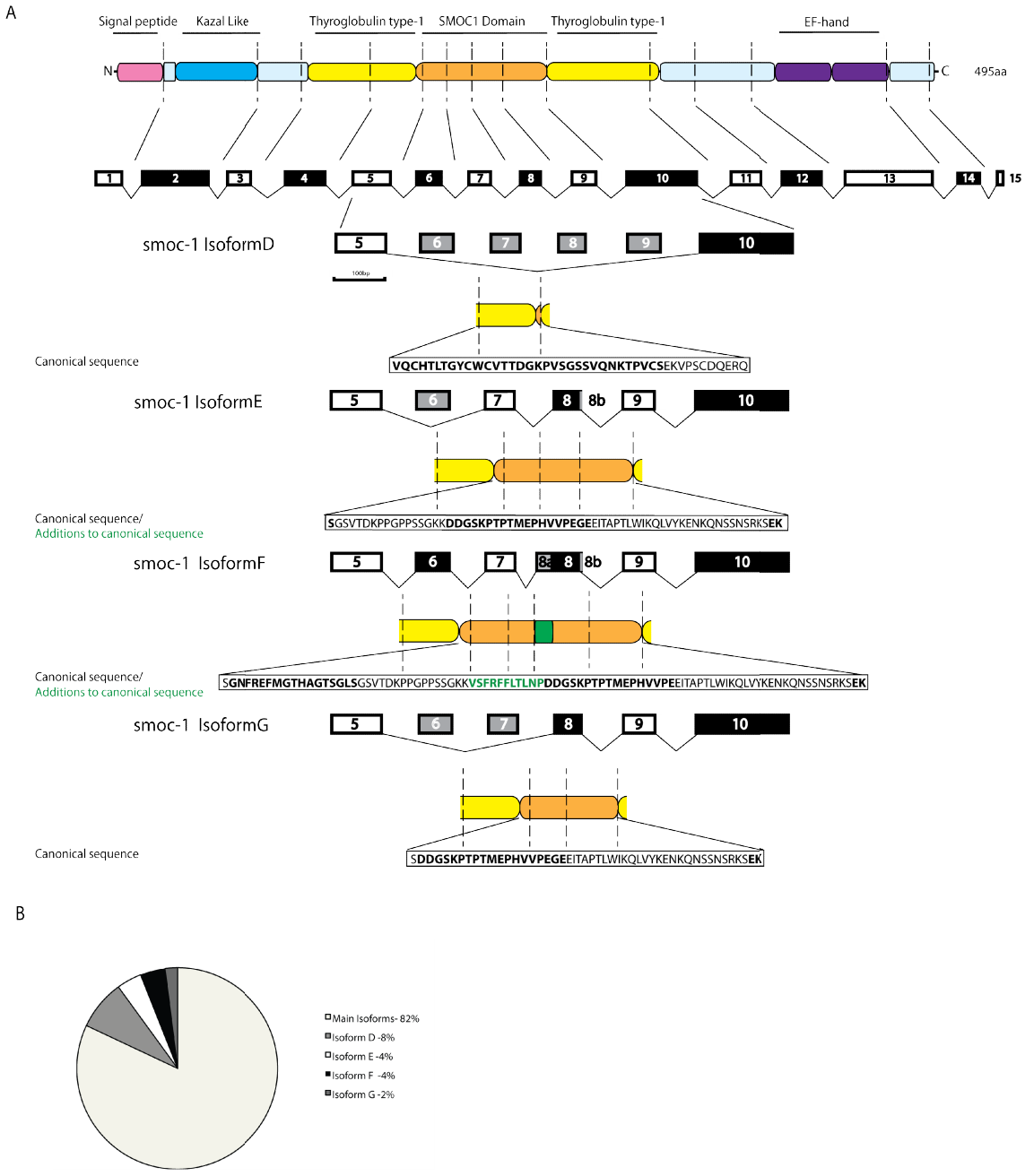



**Fig. 2.2.5 The structure of the zebrafish *smoc1* gene and its minority isoforms.**

**(A)** A schematic representation of the consensus transcript and the regions of the protein that each exon codes for. Splicing is confined to the SMOC domain with any alternatively spliced exons depicted in grey. Isoform D does not include exons 6-10 which removes 237bp (79aa) from the mature transcript resulting in the complete ablation of the smoc1 domain. Isoform E skips exon 6. It also uses the 8b splice donor site resulting in the loss of the six bp coding for the EG amino acids from the protein structure. Isoform F is spliced at both the 8a splice acceptor and 8b splice donor sites resulting in the inclusion of the VSFRFFLTLNP cassette but the loss of the EG motif from transcript. Isoform G losses exon 6 and 7 resulting in deletion of 96 base pairs (32amino acids). **(B)** The relative abundance of minor *smoc1* isoforms measured by clonal analysis. The most abundant minor splice acceptor transcript will this isoform D (8%). Followed by isoform E and F (4%). Isoform F was the least abundant identified transcript (2%).

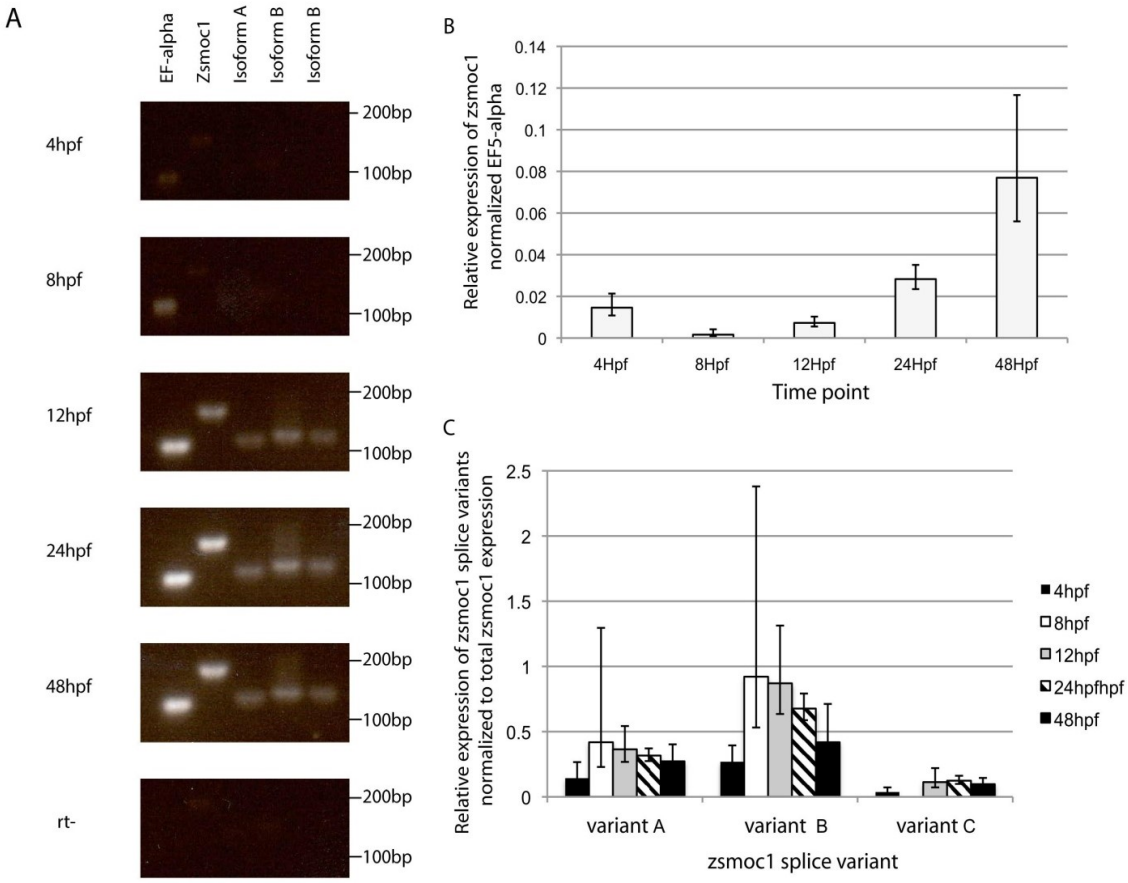



**Fig. 2.2.6 The temporal expression patterns of the major *zsmoc1* variants in early development.**

**(A)** RT-PCR analysis of *zsmoc1* expression in WT zebrafish at 4, 8, 12, 24 and 48hpf. **(B)**Relative quantification of total *zsmoc1* levels over the first few days of development via qPCR analysis. *zsmoc1* levels were calculated using the comparative Ct method and normalised to the reference gene *elfa*, N=3 error bars represent one s.e.m. **(C)** Relative quantification of the expression levels of the major variant transcripts of *zsmoc1* (variants A, B and C) *zsmoc1* levels were calculated using the comparative Ct method and normalised to the total *zsmoc1* level, N=3 error bars represent one s.e.m.

## 2.2.5 The temporal expression pattern of *zsmoc1* transcript variants

To address whether the major splice variants were developmentally regulated the expression patterns of the major splice variants were measured over the course of zebrafish development, by performing both RT-PCR **(Fig. 2.2.6 A)** and qPCR **(Fig. 2.2.6 B,C)** over the first two day of development (4, 8, 12, 24 and 48 hours post fertilisation (hpf). One pair of PCR primers was designed that would amplify every known *zsmoc1* variant in addition to primers that would only amplify specific *zsmoc1* variants. All samples were normalised to reference gene *elongation factor alpha* (*elfa*): which has already been shown to be stably expressed during early zebrafish development (Tang et al. 2007). The results of the qPCR were averaged between two biological replicates. With the notable exception of at 8 hpf, the relative expression of *zsmoc1* increases over the course of development, showing the greatest relative expression levels at the last time point measured, 48 hpf **(Fig. 2.2.6 B).** All the major splice forms were detected at every stage of development tested, with the exception of splice form C which was not present at detectable levels of sample at 8 hpf. At all stages of development splice form B is most abundant followed by splice form A and then splice form C. There is a general trend towards lower expression levels of the variants overtime, possibly because of an upregulation of the minor splice variants. However the general pattern remains the same. So there is no compelling evidence that the splice forms are regulated over the course of development.

## 2.2.6 Identification of candidate novel splice sites in human *SMOC1*

The identification of novel splice isoforms may be important for the interpretations of diagnostic sequencing in human disease. Pathogenic mutations may be missed if they are located in unknown isoforms. We have one family with definite OAS that plausibly maps to the *SMOC1* locus but in which we have not found a pathogenic *SMOC1* mutations(personal correspondence Joe Rainger 2012). This raises the possibility that the pathogenic mutation could be located in an unknown exon.

It may be difficult to identify splice isoforms in humans, particularly if they are often not very abundant, or are expressed only at specific developmental stages. It has been estimated that 95% of human genes are alternatively spliced. If this is true then between 88,000–132,000 splicing events still remain to be captured (Pan et al. 2008).

I decided to try to identify alternative transcripts in human *SMOC1* using the data I had generated in the zebrafish for cross species analysis. Using the search strategy outlined in **Fig. 2.2.7** we investigated our six novel splice variants to establish if any were conserved in humans. Zebrafish and mammals are only distantly related so the initial searches were performed using protein sequence similarity with either a blastx tblastn or tblastx.

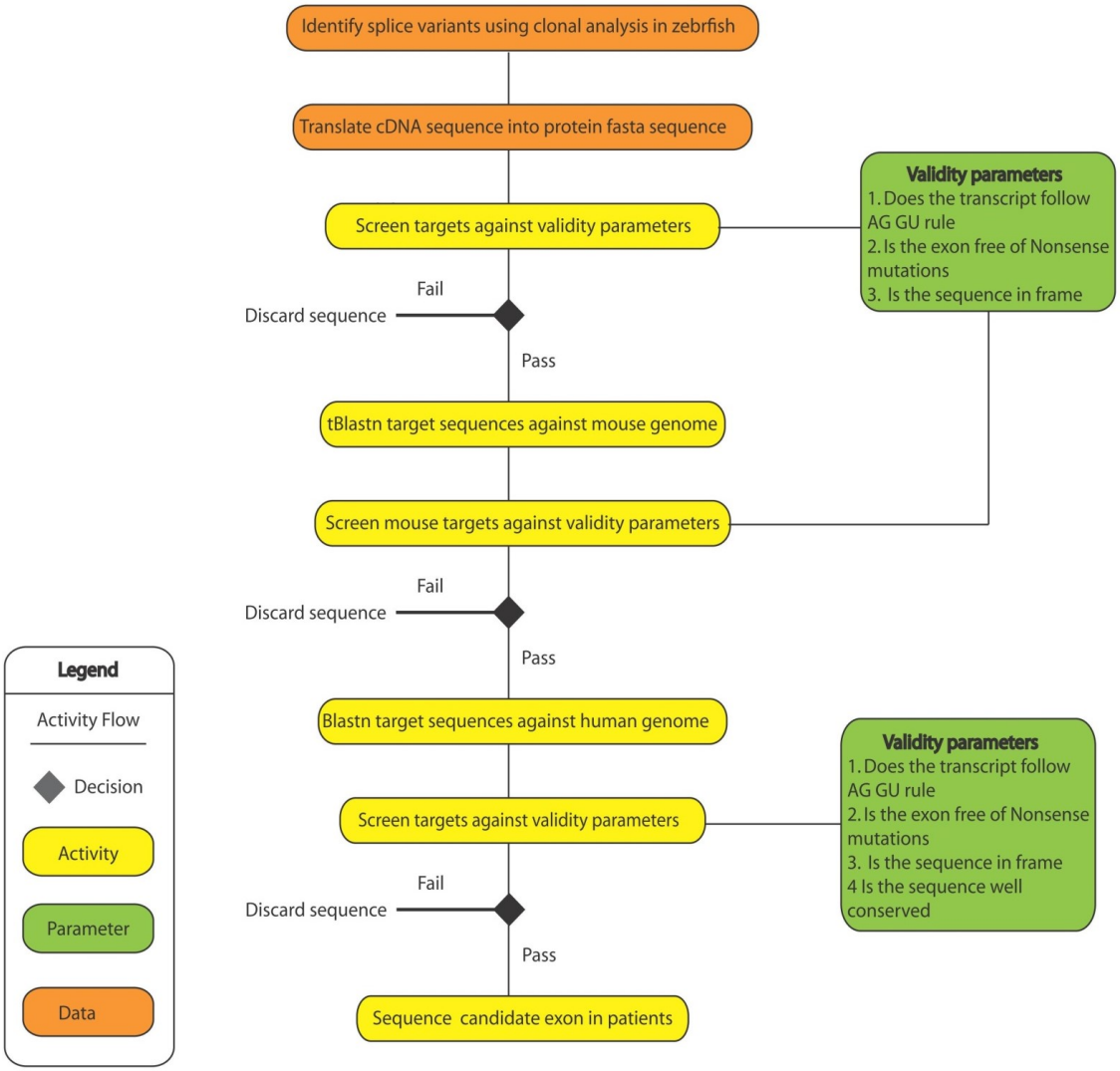



**Fig. 2.2.7 An activity flow diagram detailing the search strategy for the cross species analysis of *zsmoc1* and *hSMOC1* alternative splicing.**

Diagram details the search strategy undertaken. Initial transcripts identified from zebrafish cDNA. Splice transcripts were analysed using validity parameters. Those sequences which fulfilled every parameter moved on to the next stage of the process.

---

Initial searches yielded no hits against the human databases, possibly due to the low representation of ESTs and mRNA transcripts from early stages in human development. To address this, we searched again against the mouse databases. Of the six splice variants only one was detected in the mouse genome, this was an alternative splice acceptor site equivalent to site 8a zebrafish splice

variant C. Taking advantage of the higher conservation between human and mouse genomes it was then possible to directly query the human genomic sequence using a nucleotide query. Using this approach, a strong candidate alternative splice acceptor site was identified. Located upstream of *hSMOC1* exon6 it was almost completely conserved at the nucleotide sequence **(Fig. 2.2.8. A)**. Out of a total of 33 functional base pairs there is only a single synonymous change between mice and humans.

A multiple alignment shows that there is 68% identity between zsmoc-1 and hSMOC-1 at a protein level **(Fig. 2.2.8. B)**, with conservation of all the known functional domains with the exception of the signal peptide in zebrafish. The zebrafish Variant C sequence has two regions which are apparently unique to zebrafish marked here as U Z (unique zebrafish). The first was located between the Kazal domain and first thyroglobulin domain. This represents a "novel" exon. The second is flanked by the first thyroglobulin domain and the SMOC domain. This is the sequence attributable to the alternatively spliced exon 6 and is absent is *zsmoc1* splice variant B. The alternative splice form 8a protein sequence is conserved from zebrafish to mice with a 91% identity. If the transcript is indeed expressed in humans it would show the same level of conservation (predicted isoform 3 in **Fig. 2.2.8)**.


**Fig. 2.2.8 Multiple sequence alignments of *Smoc1* across four species.**

**(A)** A sequence alignment of human and mouse genomic sequence of *SMOC1* exon 6. Showing the canonical exon 6(blue), alternatively spliced exon 6'2 (green) and splice acceptor sequence(red). **(B)** structural alignment of translated SMOC-1 in Human, mouse and zebrafish. Total zebrafish identity is 69%. The conserved functional domains are annotated (Kazal Domain, Thryroglobulin 1 and 2, SMOC and EF hand domains) there are also regions of sequence that are not conserved in any other known animals which have been labeled UZ (unique zebrafish). The predicted alternative splice sequence in human variant 3 (exon 6') is 91% conserved in zebrafish (exon 8a) with only one amino acid being changed (red box).

All alignments were performed in ClustalW, "*****" = completely conserved sequence, "**:**" = a substitution with a highly similar amino acid, "**.**"**=** a substitution with a weakly similar amino acid. Spaces were introduced to produce the best alignments.

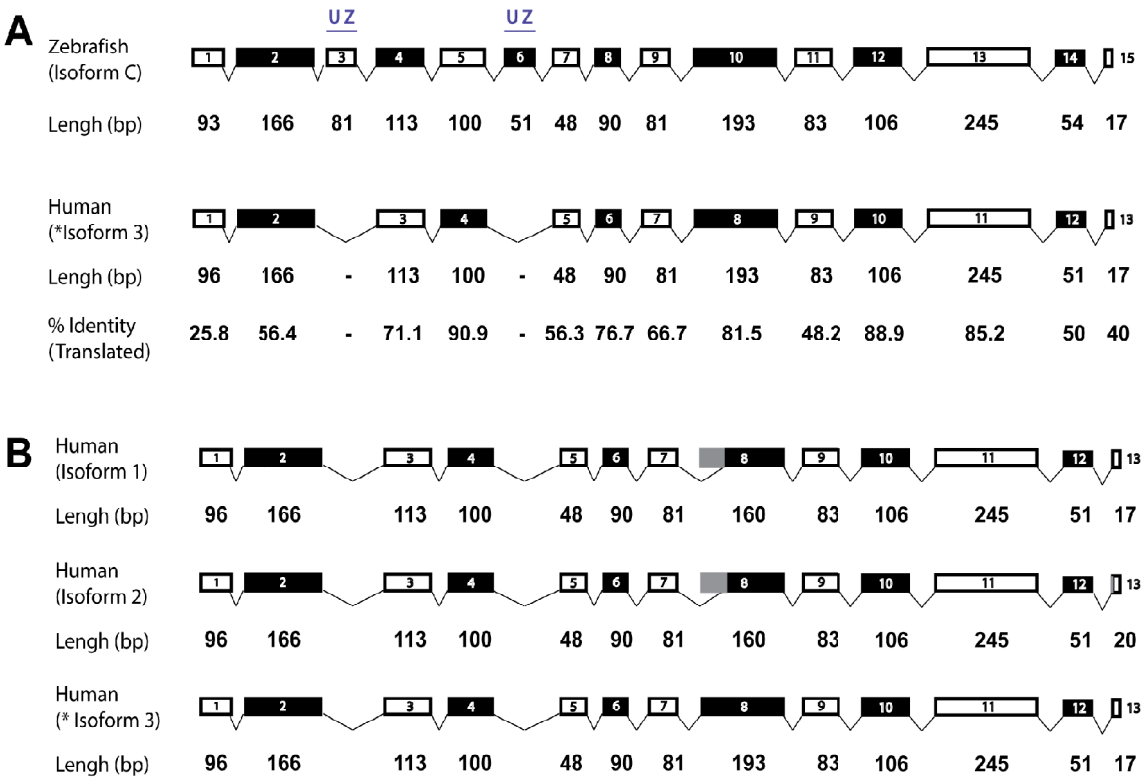



**Fig. 2.2.9 Splicing paterns are conserverd between human SMOC1 and zebrafish smoc1.**

(A) A schematic representation of the consensus transcripts of zebrafish smoc1 and human SMOC1. Intron exon boundries are completly conserved with the exception of unique zebrafish exons 3 and 6 and the splicophorm identified on exon 10 (human exon 8) (B) Schematic representation of splicing paterns of know humn isoforms of SMOC1 in comparision predicted isoform 3.

* = Predicted Isoform C

---

Analysis of the the splicing paterns of *zsmoc1* and *SMOC1* revealed that splicing paterns were indeed conserved between humans and zebrafish. The only exceptions being the unique zebrafish exon 3 and 6 **(Fig. 2.2.9 A)**. The strong consowation of splicing suports to the hypothsis that predicted isoform 3 is expressed in humans.

## 2.2.7 Functional predictions using *zsmoc1* isoforms

To help ascertain the potential function of the *zsmoc1* isoforms a bioinformatics approach was used to analyse protein structure in silico. Initially the *zsmoc1* isoforms A B, and C were analysed using PONDR-VLXT (Li et al. 1999;Romero et al. 2001; Romero et al. 1997) to see if regions of alternative splicing corresponded with intrinsically disordered regions (IDRs) within zsmoc1. IDRs are the domains of a protein with no defined three-dimensional structure(Dunker et al. 2008; Dyson & Wright 2005). Which means the amino acids making up an IDR have complete freedom of movement in vivo. It has already been established that there is a bias for alternatively spliced exons to be located in the disordered regions in proteins, this is particularly true for symmetrical exons (Romero et al. 2006). The largest predicted region of disorder in zsmoc-1 is in isoform B **(Fig. 2.2.10 B)** and is located between residues 143 and 200 of the SMOC domain. This is the region in which all of the alternative splicing events are localized **(Fig. 2.2.4; Fig. 2.2.5).** This is the only statistically significant predicted IDR in zsmoc1. The alternative exon 6 in isoforms A and C and splice acceptor site 8a are both predicted to be highly ordered motifs within this region of disorder **(Fig. 2.2.10 B and C)**. This implies that they are functional motifs and do not have a purely structural function.

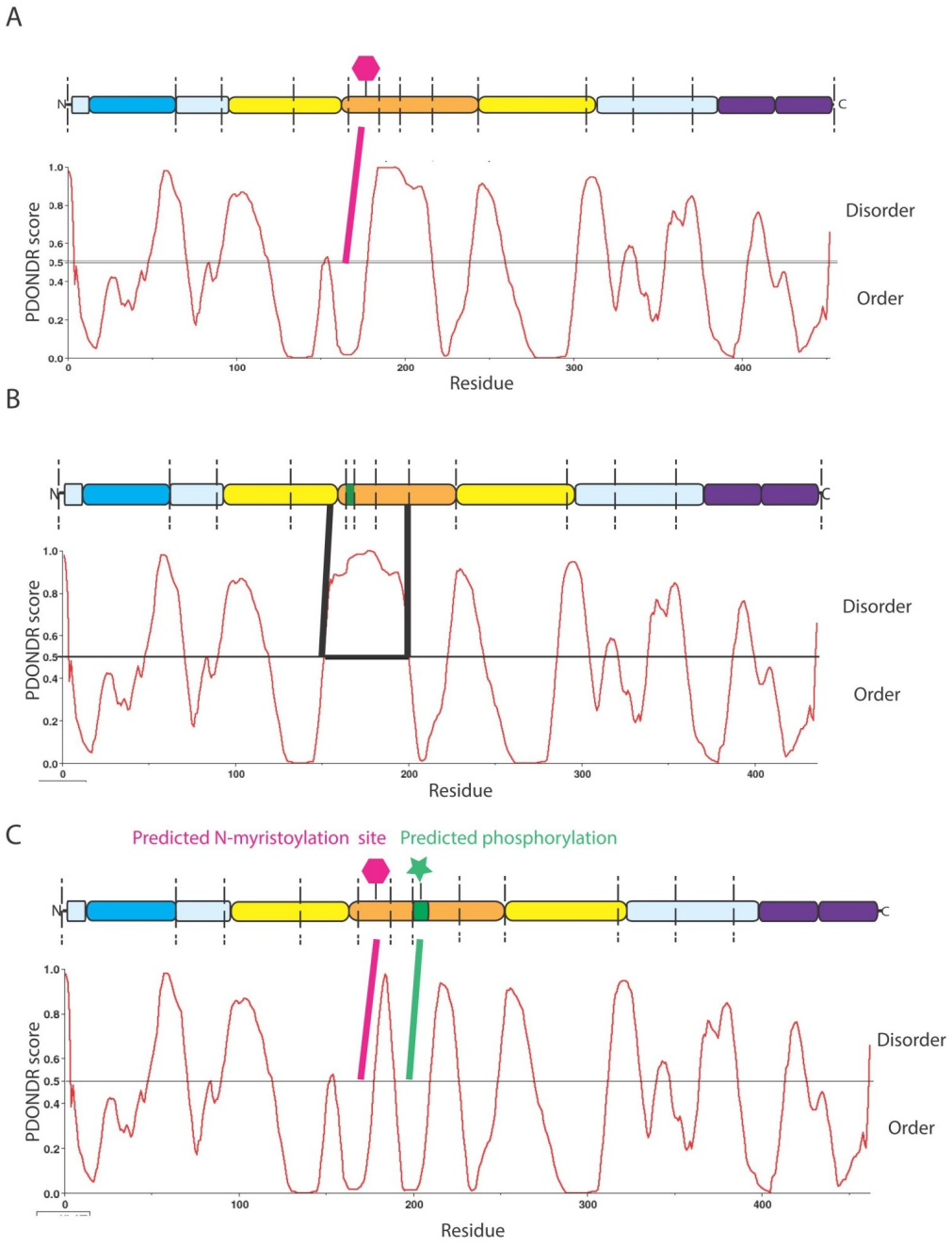



**Fig. 2.2.10 In silico analysis of the structure of zsmoc-1 for the alternative splice variants.**

Predicted intrinsic disorder of per residue of zsmoc-1 splice variant A **(A)**, B **(B)** and C **(C)** against a schematic representation of the translated protein. The predicted disorder of each residue in the translated protein was scored by the PDONDR VT by Matthews correlation. Regions with a Matthews correlation of greater than 0.5 were predicted to be disordered. Regions that were predicted to be disordered for 40 consecutive amino acids were deemed significant. The region highlighted in splice variant B was the only significantly disordered region. Post translational modifications predicted by Motif scan are highlighted in pink for N-myristoylation and green for phosphorylation.

The predictions were made by Motif scan (2013) of a single post-translational modification in each alternatively spliced region of the major zsmoc-1 isoforms provide support this hypothesis. Motif scan predicted a N-myristoylation sites in *zsmoc1* exon 6 and an additional phosphorylation site exon 8a **(Fig. 2.2.9. C).**

# 2.3 Discussion

## 2.3.1 The zebrafish genomic locus has been established

The misleading, fragmented genomic annotation of *zsmoc1* hampered my efforts to understand this protein. I have demonstrated that *zsmoc1* is present as one contiguous gene in zebrafish. The current annotation also misses a functionally important 5' section of the open reading frame. Although the zebrafish genome was supposedly completed in 2013(Howe et al. 2013) the *zsmoc1* locus is still misannotated in the public genome browsers. This could be due to the fact a large part of the genomic sequence still relies on shotgun sequencing. Currently 17% of the current build of the zebrafish genome is aligned from shotgun sequencing (Howe et al. 2013).

The missing *zsmoc1* signal peptide has been found**(Fig. 2.2.3)**. The coding sequence of the first exon is almost completely comprised of the open reading frame coding for the signalling peptide in zebrafish and indeed in every vertebrate species and in *Drosophila Pent* **(Fig. 2.2.2; Fig. 2.2.3)**. This arrangement creates a problem for automated gene analysis software because signalling peptides are not the most strongly conserved domains. It is the overall hydrophobicity of the amino acid sequence rather than the specific amino acid order that targets the signal peptidase to a signal peptide for cleavage **(Fig. 2.2.2;** Petersen et al. 2011**).** The

signal peptide in zebrafish is spliced over a large first intron again similar to the situation in other vertebrates where the first intron is between 50 -80 kb away (**Sup. Table 1)**.

The intron-exon boundaries and splice arrangement of the of the first exons of the *Smocs* are remarkably well-conserved across evolution **(Fig. 2.2.2; Fig. 2.2.3)**.

## 2.3.2 Alternative splicing in *zsmoc1* action

Alternative splicing is essential to understanding the eukaryotic proteome. First discovered in the late 70s (Berget et al. 1977; Chow et al. 1977) it rapidly became apparent that the vast majority of eukaryotic transcripts are spliced (Kelemen et al. 2013). This largely explains some of the discrepancy between gene number and protein number in eukaryotes. *zsmoc1* is no exception **(Fig. 2.2.4; Fig. 2.2.5).** I have identified seven separate splice transcripts of *zsmoc1*, six of which are novel and have not been reported. All identified splicing was confined to between exons 6 and 10; between the thyroglobulin domains in the region identified as the SMOC domain raising the question of whether the splice variants have any different functionality. RUST(regulated unproductive splicing and translation) is a mechanism to control the level of mature transcript by introducing frame shifts into the coding sequence and trigger nonsense mediated decay. RUST accounts for a large fraction (25 to 35%) of all alternative splicing

events (Lewis et al. 2003). RUST can be convincingly ruled out in *zsmoc1* as all alternatively spliced variants I have identified are symmetrical (i.e. the spliced region has a base pair length which is a multiple of three so will usually preserve the reading frame of the transcripts whether the exon spliced or not; Magen & Ast 2005). This places our alternative transcripts in the other 65%-75% of transcripts that have altered transcripts within the coding region.

There are a wide range of known functional possibilities. Alternative splicing has been known to change the stability (Sakurai et al. 2001), localisation (Prou et al. 2001) and catalytic activity (Li & Koromilas 2001) of proteins. In silico analysis on isoforms A, B and C found splicing was confined to the largest region of predicted intrinsic disorder within zsmoc-1 **(Fig. 2.2.9).** This is in line with predictions that alternative splicing is preferentially found within intrinsically disordered regions (IDRs) as they are less structurally constrained (Wang et al. 2005; Romero et al. 2006). A change within an allosteric region of a protein will disrupt that structure resulting in loss of function. IDRs have no defined structure to disrupt, which means alternatively spliced exons in these regions are much less likely to disrupt the overall structure of a protein. IDRs in proteins have several known functions, including providing structural flexibility to proteins by allowing them to move dynamically and are often also associated with protein binding sites where they will only form a defined structured when bound to their ligand. (Pancsa & Fuxreiter 2012). Intriguingly, the alternative splice sites in the less abundant splice variants A and C result in a dramatic reduction in predicted disorder, suggesting they are functional motifs, and their short amino acid

composition implies they are post translational modification sites. Motif scan predictions supported this, with post translational modification predicted in both splice exons 6 and 8a **(Fig. 2.2.9.).** This would not be a surprising, as the introduction of post translational modification sites is the most common identified function of alternative splicing (Kelemen et al. 2013b).

There is a danger in over interpreting in silico predictions without validation by experimental data. If the relative expression of the splice variants had changed with time then we would have shown that the variants were developmentally regulated. Yet, the ratios remain consistent **(Fig. 2.2.6)**, so while a functional role of the splicing variants cannot be ruled out neither can it be ruled out that they are simply co-expressed exons with no differential function. The alternatively spliced exons are too short for in situ hybridisation so that at the present time it is impossible to establish whether they have different localisations during zebrafish development **(Fig. 2.2.4; Fig. 2.2.5)**.

Given the difficulty of obtaining direct functional evidence for different splice variants most are analysed purely on the basis of their evolutionary conservation. If a splice site is conserved then presumably it must be functionally relevant otherwise there will be no purifying selection to maintain it. Under this analysis splice variant C is likely to be a functional variant as it shows high conservation from mammals to zebrafish (91%**; Fig. 2.2.7. B).**

## 2.3.3 Cross species analysis as a means of detecting alternative transcripts in humans

While cataloging the alternative splicing patterns in genes from different species is interesting from an evolutionary standpoint, it is also potentially medically relevant; the majority of hereditary diseases are thought to be due to aberrant splicing (Ward & Cooper 2010). Most documented cases are caused by mutations directly to splice donor (GT) or splice acceptor sites (AG) (Krawczak et al. 2007). The sequences are essential for the correct splicing, and their mutation can lead to a whole range of defects. An exons can be excluded, introns included and cryptic splice sites can be activated within genes. These mutations are often functionally nulls. In two thirds of cases aberrant splicing results in frame shifts which, along with intron inclusion, leads to premature stop codons and NMD (Magen & Ast 2005). As is the case for the mutation of *SMOC1* identified in Abouzeid et al. (2011). The patient had inherited a homozygous splice donor site mutation in exon 4 which led to OAS. Two further SMOC1 mutations were identified with splice donor site mutations in Okada et al. (2011) which means 25% of the known cases of OAS are caused by missplicing. As patients with splice mutations display the same severity of both anophthalmia and limb abnormalities it is assumed that these defects are functionally nulls rather than hypomorphs. Even an alteration in the ratios of spliceoforms can profoundly affect the phenotype of the disease. Mutations in the splice sites of *ATP7A* inhibit copper absorption, leading to Menkes disease (MD). Yet mutations to less strongly

conserved bases in the same splice sites lead to the much less severe occipital horn syndrome (OHS) (Møller et al. 2000; Das et al. 1995).

Unidentified exons can also harbor mutations that lead to pathology. These would then be missed when screening patients. For example in 1998 it was discovered that mutations in *USH2A* lead to the development of syndromic deaf blindness, Usher syndrome type IIa (Eudy et al. 1998). In the following six years mutations were identified in *USH2A* in 63% of Usher type IIa patients. That still left 37% of patients without an identifiable mutation. Wijk et al. identified 51 novel exons for *USH2A,* and went on to identify mutations in five out of 12 patients located in these exons (van Wijk et al. 2004). This accounted for ~20% of Usher type IIa patients, leaving only ~17% of cases unexplained. Exon skipping can also rescue nonsense mutations, which would lead to less severe symptoms. Mutations in *Dystrophin* provide a good example of this. It is a large protein (427 kd) with a correspondingly large genome locus made up of 79 exons. Mutations in *Dystrophin* lead to 2 related syndromes: Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD). Both syndromes are usually caused by indels (80% ). Deletions that are out of frame are functionally nulls and lead to DMD, while in frame deletions lead to the milder BMD (Muntoni et al. 2003). To help study these diseases mutations in *Dmd* have been induced to produce the *mdx* mouse which mimic the features of the disease, the main one being a 60-70% reduction in type A γ-aminobutyric acid ($GABA_A$) receptor clusters. Exon skipping has been artificially induced in *mdx* mouse. These skipped exons restore the reading frame of the *Dma* gene and greatly reduce the DMD phenotype including completely rescuing $GABA_A$ cluster numbers(Vaillend et al. 2010).

The cataloguing of alternative splicing events in the human genome remains incomplete with estimates ranging from 40 to 50% of human splice variants still to be uncovered. This is because of the drawbacks of EST technology. While ESTs provide powerful sequence depth, quality remains low. EST sequences are based off single reads and so often include sequencing errors and chimeric cDNA. While there is some debate, the splicing patterns of alternatively spliced genes are thought to be less well conserved than those of non-spliced genes(Modrek & Lee 2003; Kriventseva et al. 2003). Still functionally important splice forms should be conserved and these account for the most interesting splice sites. The comparison of splice variants from different species has been used to perform cross species analysis and successfully identify unannotated splice forms within different genomes (Lu et al. 2010; Chen et al. 2006). The availability of novel *zsmoc1* splice variants allowed the performance of a cross species analysis which led to the identification of a potentially novel human SMOC1 splice variant **(Fig. 2.2.7)**.

Mutations have not been identified in 42% of our patients with OAS. This raised the possibility that as in the case of *USH2A* these mutations were hidden in this previously unidentified exons (van Wijk et al. 2004). With this in mind the cohort of OAS patients where the causative mutation had not been identified were screened for mutations by Joe Rainger (Personal correspondence 2013). However, no mutations were identified.

## 2.3.4 Concluding remarks

In this chapter I have established that zebrafish contains at least one full-length orthologue of *SMOC1*. The orthologue is developmentally regulated and it has multiple previously unrecognized spliceoforms. I have also demonstrated that cross species analysis can be used to identify potentially novel human exons, which could be of clinical relevance.

# Chapter 3

# zsmoc-1: an antagonist of bmp signalling

# 3.1 Introduction

## 3.1.1 Developmental signalling

The central question of developmental biology is how a single cell with a generic identity generates a whole organism comprised of multiple organs, which in turn is comprised of different tissues. Tissues themselves are mixtures of different cell populations. The development of an animal requires the regulation of cell fate, pluripotency, proliferation, apoptosis and migration. How is this specificity of cell identity achieved? In *Caenorhabditis elegans* the specification of cell fate is accomplished almost entirely by cell division(Sulston & Horvitz 1977; Kimble & Hirsh 1979; Fay 2005), however, this is unusual.

In most organisms developmental specification of cell identity is achieved via secreted ligands: secreted molecules (usually proteins) that can be detected by receptors most commonly sited on the surface of a cell. There are multiple signaling pathways, though the most significant for development are the tgf-beta Shh, Mapk and Wnt pathways (Pires-daSilva & Sommer 2003). While each family is involved in multiple processes and there is considerable evidence for cross talk. This chapter is focused on the BMP signalling pathway(Verheyen 2007), as this is the pathway that both the work presented and the existing literature links to SMOC-1(Thomas et al. 2009, p.1; Vuilleumier et al. 2010; Hamaratoglu et al. 2011).

## 3.1.2 The Bmp signalling cascade

Bone morphogenic proteins (Bmp) are a sub family of the TGF-beta superfamily. The major components of the Bmp cascade used during development have been conserved from *Cnidaria*, *Drosophila* to humans. The elements of this signalling cascade are well studied. There are 11 different Bmp ligands in the Bmp subfamily. These are split into four different subfamilies which are involved in both distinct and overlapping functions (Miyazono et al. 2010). For example, while bmp2b is involved in the dorsoventral patterning of the zebrafish embryo during gastrulation (Martnez-Barber et al. 1997), it also specifies the dorsal identity of the zebrafish retina (Kruse-Bend et al. 2012). While bmp7 (Little & Mullins 2009) and gdf6 are exclusively involved in embryonic and retinal patterning (Gosse & Baier 2009) respectively, some of this is due to the dimeric nature of bmp ligand. Both homomeric and heteromeric dimers are expressed, but they are not functionally redundant. In zebrafish embryonic development, it is only heteromeric bmp dimers that are functional. Both bmp2b/7 and bmp2b/4 have dorsalising activity, but homodimers do not, which is demonstrated by the failure of homodimeric bmp2b to rescue *bmp2b* nulls (*bmp2b* morphants) while heterodimeric bmp 2/7 could (Little & Mullins 2009).

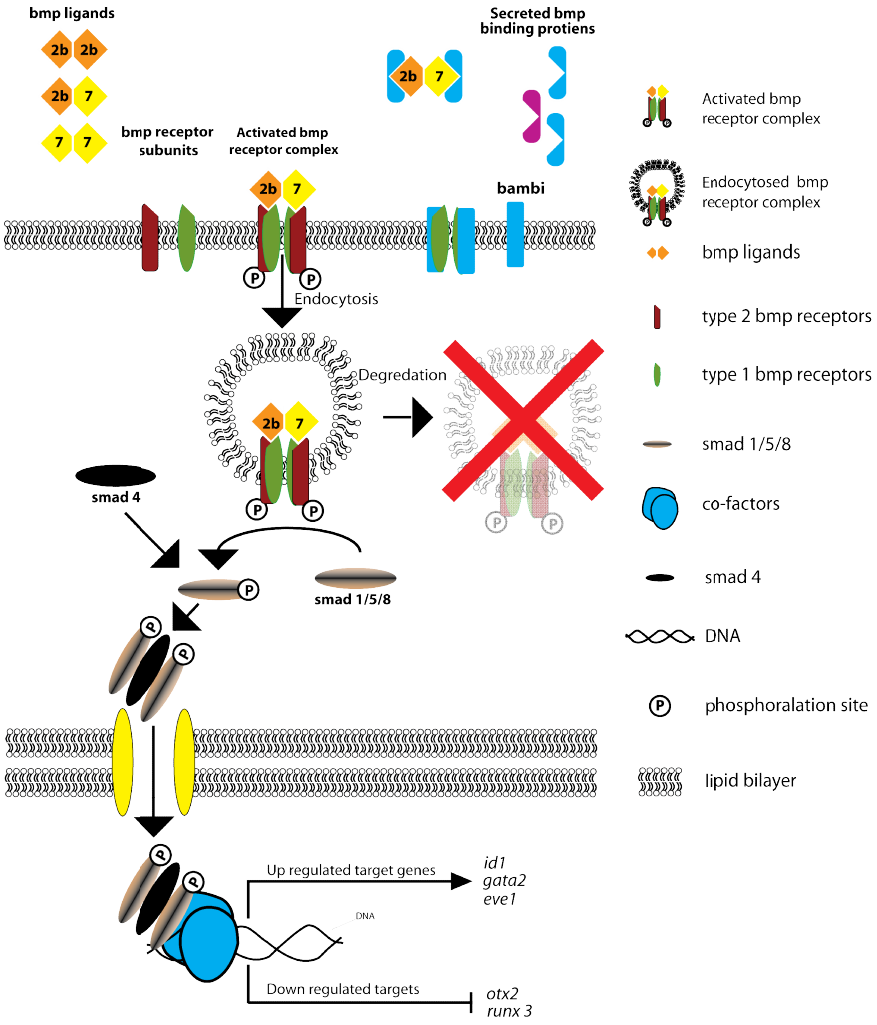



**Fig. 3.1.1 The canonical in bmp signalling pathway.**

A schematic representation of the core elements of the bmp signalling pathway activated during the dorsoventral patterning of the zebrafish embryo. Signals initially carried in the form of an extracellular gradient of bmp ligand. The ligand binds to the bmp receptor complex which triggers it to auto-phosphorylate. Once the receptor is phosphorylated, it is endocytosed, Whereupon it can activate a co-smad (1/5/8). The activated co-smad forms a trimeric complex with smad 4 before accumulating in the nuclease where it acts as a transcription factor.
The figure also illustrates that bmp cascade is regulated at several levels. Secreted bmp binding proteins can prevent ligand binding to the bmp receptor complex while pseudo receptors can prevent the formation of activated bmp complex entirely. The pool of active components is continually regulated by the degradation endocytosed ligand receptor complexes.
(Data reviewed in Ramel & Hill 2012 and Schwank et al. 2011).

The various bmp ligands are recognised by bmp receptor complex, which is a membrane-bound tetramer. Broadly there are two types of sub receptors, type 1 and type 2. Bmp ligands preferentially bind to the type 1 receptors, which then recruit the type 2 receptor. The Type 2 receptor activates the complex by phosphorylating the receptors GS domain. There are multiple types of each receptor (Ehrlich et al. 2011), which allows the formation of a variety of combinations of receptor complex, each with unique binding affinities. A good example of this is zebrafish dorsoventral patterning. Zebrafish have at least five known type 1 receptors (alk3a, alk3b, alk6a, alk6b and alk8; **table 3.1**) and two known type 2 receptors (bmpr2a and bmpr2b; **table 3.1**). Little & Mullins used morpholinos to knock-down each type 1 bmp receptor. *Alk8* MO knockdown resulted in a complete ablation of the phospho-smad 1/5/8 (hereafter termed p-smad) gradient. In contrast individual *alk3a, alk3b, alk6a* and *alk6b* MO knockdowns have no phenotype (or very mild dorsalisation), demonstrating that they are functionally redundant. However, when *alk3a, alk3b, alk6a* and *alk6b* are all knocked down simultaneously, the p-smad gradient is ablated (Little & Mullins 2009). This demonstrates that the functional bmp2r complex is a heterodimeric complex comprised of an alk8 protein and one of either alk3 or alk6.

| Pathway | BMP ligand | Type 2 receptors | Type I receptors | R-smad | Co-smad |
|---|---|---|---|---|---|
| zebrafish embryo dorsoventral patterning | **bmp2b**<br>bmp4<br>**bmp7a** | **bmpr2a**<br>**bmpr2b** | **alk3a**<br>**alk3b**<br>**alk6a**<br>**alk6b**<br>**alk8** | **smad1**<br>**smad5**<br>**smad8** | **smad4** |
| Zebrafish retina dorsal ventral patterning | **GDF-6**<br>***bmp2b*** | bmpr2a<br>bmpr2b | alk3a<br>alk3b<br>alk6a<br>alk6b<br>alk8 | **smad1**<br>**smad5**<br>**smad8** | **smad4** |
| *Drosophila* wing patterning | **Decapentaplegic**<br>Screw<br>Glass Bottom Boat | wishful thinking | **Thickveins**<br>Saxophone<br>Punt | **MAD** | Medea |

**Table 3.1.1 The components of the core bmp signalling pathway.**
Zebrafish proteins are denoted in blue while *Drosophila* proteins are in purple. Proteins that are only predicted to be involved in the processes are depicted with grey fonts. Proteins mentioned in the text are highlighted with bold fonts.
Table compiled from information obtained in (Ramel & Hill 2012)(Miyazono et al. 2010)(Gosse & Baier 2009)(Kruse-Bend et al. 2012).

---

Once the receptor complex is activated it phosphorylates regulatory smads (r-smads). There are three r-smads in vertebrates (smad 1/5/8; **Table 3.1.1)**, and they share extremely conserved amino acid sequences (Arnold et al. 2006). There is a degree of functional redundancy; for example mouse knockouts of *Smad1* nulls, *Smad5* nulls and *Smad1*$^{+/-}$:*Smad5*$^{+/-}$ double heterozygotes all die at E10.5 while displaying identical abnormalities (Arnold et al. 2006). Yet r-smads are not entirely redundant. In situ hybridisation shows that *smad1* and *smad5* have distinct expression patterns in gastrulating zebrafish embryos and these

expression patterns are under different systems of regulation. *smad1* expression is up regulated by *bmp2b,* while *smad5* is ubiquitously expressed (Dick et al. 1999). Additionally, exogenous *smad1* can rescue *bmp2b* nulls (*swirl* mutants) but *smad5* cannot (Dick et al. 1999).

Once they are phosphorylated a pair of r-smads will bind to a co-smad (smad 4 in vertebrates) forming a trimer (Kruse-Bend et al. 2012). The complex is then translocated into the nucleus, where it acts as a transcription factor. It binds to target sequences which will then alter the expression of target genes, either up-regulating or down regulating them depending on the context(Ramel & Hill 2012). In gastrulating zebrafish embryos, bmp up regulates *id1* (Collery & Link 2011), *eve1* (Seebald & Szeto 2011; Barro et al. 1995) and *gata2* (Maeno et al. 1996; Oren et al. 2005; Sykes et al. 1998) expression, whilst simultaneously down regulating *otx2* (Navaneetha Krishnan Bharathan, 2014) and *run3x* (Cohen 2009). Whether the gene is turned on or switched off by smad binding is dependent on the cofactors it binds with (Ramel & Hill 2012; Arora et al. 1995; Dai et al. 2000*;* Cheng et al. 2007).

While binding partners for the smads have been identified (e.g. HoxC8, Nkx3-2 and id1; Ramel & Hill 2012) their precise function remains uncharacterised. Of the known cofactors, Schnurri (Arora et al. 1995) is the most well studied, but like the smad complex, its effects appear to be context specific; up regulating some genes (e.g. *Brinker* in *Drosophila*; Pyrowolakis et al. 2004) while down regulating others (e.g. *Ubx B* in *Drosophila;* (Dai et al. 2000)) . Cofactors have been identified that impart tissue specificity, which accounts for the reuse of bmp signalling cascades in different tissues. Runx2 for example is the key Bmp binding

factor in bone formation (reviewed in Karsenty 2008). Knocking it down inhibits bone formation as even the presence of Bmp is essential for osteoblast differentiation. *Oaz* has been shown to mediate late onset bmp signalling with neuronal defects in both the brain (Cheng et al. 2007) and nasal cavities (Cheng & Reed 2007) reported in knockout mice. However, the knockout mice do not show any defects you would observe if Bmp signalling was disrupted early on in mouse development.

## 3.1.3 Bmp signalling gradients

Secreted bmp ligands impart spatial information to the embryo via their action as morphogens. The so-called French flag hypothesis (Wolpert 1969) posits that p-smad activity is concentration dependent and that cell fate is determined by reaching specific thresholds of p-smad activity. There is some evidence that this is the case; as would be predicted, the bmps have been observed forming the stable gradients of p-smad activity in sea urchins (Lapraz et al. 2009), *Drosophila* (Akiyama et al. 2008), *Xenopus* (Plouhinec & De Robertis 2009) and zebrafish (Gosse & Baier 2009). Only in *Drosophila* has the Dpp gradient been directly observed. GFP tagged-Dpp was used to quantify the gradient in real time, which showed that in *Drosophila* ligand concentration decreases exponentially as the ligand migrates further from the source(Teleman & Cohen 2000). This is the same pattern that would be predicted for a classical morphogen (Teleman & Cohen 2000). Additionally, the downstream up-regulation of downstream targets of Dpp

has been shown to be dosage dependent, again this is consistent with the action of a morphogen.

In the dorsoventral patterning of the zebrafish embryo up-regulation of p-smad signalling has been shown to phenocopy bmp mutants and morphants (von der Hardt et al. 2007; Topczewski et al. 2001); morpholino knockdown of chordino (an antagonist of BMP signalling) and perturbation of the bmp gradient with Knypek morpholino knockdowns both lead to defects in the lateral migration and therefore the extension movements of the gastrulating zebrafish embryo. This is unexpected and the opposite of what would be predicted if bmp does act as a morphogen. Hardt et al. (2007) went on to demonstrate that these defects are caused by the disruption of the shape of the bmp gradient in the embryos. The laterally migrating cells followed the shape of the gradient, from high to low psmad expression. In the chordino mutants, the bmp's were ubiquitously expressed so there was no gradient for the lateral cells to follow (von der Hardt et al. 2007).

Whether or not bmp is acting as a classical morphogen, successful signalling is reliant on the formation of the gradient. This gradient must be dynamically controlled and maintained. If a gradient is formed by unregulated passive diffusion and degradation of the signal would be susceptible to perturbation (for example genetic mutation, temperature and the availability of nutrients ; Eldar et al. 2003). Developmental gradients must also scale with the growth of the embryo/organ (Barkai & Ben-Zvi 2009). To achieve this, the gradient must be actively controlled and indeed decades of research have identified proteins that regulate the activity of the Bmp signalling cascade at every level **(Table 3.1.2)**.

When Bmp ligand is secreted, it is subject to the antagonism of secreted bmp binding proteins such as Chordin and Twisted gastrulation **(Table 3.1.2**; Ramel & Hill 2012). In zebrafish gastrulation these proteins are expressed dorsally, forming a gradient directed towards the ventral side of the embryo. They block bmp signalling by binding to bmp dimers and preventing them from interacting with bmp receptor proteins (Piccolo et al. 1996; Chang et al. 2001; Branam et al. 2010), antagonising the Bmp action more as the bmp gradient extends dorsally. Another secreted Bmp antagonist is Tolliod. Tolliod is a protease that binds to the inactive bmp chordin complex, cleaving the chordin and releasing the bmp and thus upregulates bmp signalling (Pappano et al. 2003).

Another form of extracellular control is regulating the shape of the gradient by controlling the speed at which the ligand travels. The availability of finely quantified gradients means that *Drosophila* has been the model organism of choice for investigating these problems. Work on the Bmp homologue Decapentaplegic (Dpp) has demonstrated that proteins exist that regulate Dpp migration. Most notably the glypican homologues Dally and Dally-like as they control the shape of the gradient are neither classical antagonists nor agonists, and have been shown to both positively and negatively regulate Dpp signalling depending on context(Akiyama et al. 2008; Ramel & Hill 2012; Belenkaya et al. 2004). What is clear is that Dpp migration is dependent on Dally, an extracellular heparan sulphate protein that is homologous to the vertebrate Glypicans. GFP-Dally cannot migrate across the *Dally* null regions of the imaginal disk (Belenkaya et al. 2004). It is also clear that Dpp signalling is dependent on endocytosis, and that inhibiting endocytosis by inhibiting Dynamin results in the complete inhibition of MAD

phosphorylation (Belenkaya et al. 2004; Entchev et al. 2000; Akiyama et al. 2008). However there remains a debate as to the precise mechanism of Dpp migration, and consequently also of Dally action.

The two most popular models are receptor-mediated transcytosis (RMT) and restricted extracellular diffusion (RED)(Schwank et al. 2011). In RMT, the receptors are endocytosised and re-secreted on the other side of the cell. Thus endocytosis is used to ferry the ligand across cells and establish the Dpp gradient. There is evidence to support this. Dpp diffusion in the wing disk is dependent on the presence of Thickveins (the *Drosophila* homologue bmpr1) and GFP-Dpp diffusion is blocked in temperature sensitive *Dynamin* mutants at non-permissive temperatures (Entchev et al. 2000). However opponents of RED point out that Dpp gradients form and spread very rapidly. Which means that endocytosis and re-secretion would have to occur within minutes. Furthermore, even when transcytosis has been observed an extracellular pool of Dpp is always also present. They argue that passive diffusion would establish a gradient with a rapidity that better fits the observed speeds of Dpp migration (Schwank et al. 2011; Kicheva & González-Gaitán 2008). In RED the endocytosis targets both the ligand and receptors necessary to both smad phosphorylation (which in this model is confined to endocytosis these vesicles) and for ligand/receptor degradation, which in turn lowers the available pool for gradient formation. Of course both models are not mutually exclusive.

It remains to be investigated whether endocytosis is important in gradient formation in bmp signalling in zebrafish, or even in vertebrates. However, signalling endosomes are well documented in other TGF-beta signalling pathways

(Di Guglielmo et al. 2003; Mitchell et al. 2004), Endocytosis of BMP has been observed in the multiple cell lines (C2C12, 293T, and COS7) with different endocytotic vesicles having distinct SMAD phosphorylating activities (Hartung et al. 2006; Akiyama et al. 2008). Additionally, heparin sulphate proteins like Dally have been shown to restrict bmp gradients, and when heparin sulphate groups are digested by artificially expressing *heparitinase I* mRNA in *Xenopus* bmp expression becomes ubiquitous (Ohkawara et al. 2002).

Another level of control of the signal gradient is on the formation of the bmp receptor complex. The bmp pseudo-receptor bambi (also known as radar in zebrafish) can also bind to type 1 bmp receptors, as bambi has no phosphorylating activity and any receptor complexes it forms are inactive(Ramel & Hill 2012). This makes Bambi a classical antagonist of bmp signalling. An example of this is the action of bambi during zebrafish gastrulation, where it promotes the dorsal cell fate (Gosse & Baier 2009).

Once the r-smads are phosphorylated they can be regulated intracellularly by inhibitory smads (i-smads; smads 6/7). The inhibitors smad 6 and 7 repress bmp signalling via a number of mechanisms. Both smad6 and smad 7 compete with R-smads by binding directly to type 1 bmp receptors (Imamura et al. 1997; Hayashi et al. 1997). Additionally, smad7 recruits E3 ubiquitin ligases to the receptor complex, thereby upregulating their degradation (Kavsak et al. 2000), while smad6 competes with the R-smads for binding to smad4 (Hata et al. 1998). They are both expressed in zebrafish during gastrulation and presumably antagonise bmp signalling, though this remains to be confirmed (Hammerschmidt & Mullins 2002).

| Modulators | Homologue | Type | Effect |
|---|---|---|---|
| knypek | Glypican 4 | Heparan sulphate proteoglycan | Positive |
| **chordin** | | Secreted BMP binding protein | Negative |
| **Twisted** | | Secreted BMP/Chd binding protein | Positive/negative |
| **Tolloid** | | Protease | Positive |
| **bambia** | | BMP pseudo-receptor | Negative |
| bambib | | BMP pseudo-receptor | Negative |
| | | | |
| **Dally** | Glypican | Heparan sulphate proteoglycan | Positive/negative |
| **Dally-like** | Glypican | Heparan sulphate proteoglycan | Positive/negative |
| **Pentagone** | smoc-1 | | Positive/negative |
| Sog | Chordin | Secreted BMP binding protein | Negative |
| **Schnurri** | Schnurri | Transcription factor | Positive |
| **Brinker** | None | Transcription factor | Negative |

**Table 3.1.2 Known modulators of bmp signalling**
Zebrafish proteins are denoted in blue while *Drosophila* proteins are in purple. Proteins mentioned in the text are highlighted with bold fonts.

Table compiled from information obtained in Ramel & Hill (2012) and Miyazono et al. (2010).

The last known level of control occurs at the DNA binding level. Certain transcription factors block the binding of smad trimers to their target sequences, most notably Brinker in *Drosophila*, which is known to block all Dpp activity in the developing wing disk (Blitz & Cho 2009).

## 3.1.4 bmp gradients in zebrafish development: dorsoventral patterning of the zebrafish embryo and eye

While bmp gradients are utilised in multiple separate processes in zebrafish development, the best studied system in zebrafish is the dorsoventral gradient formation over the course of gastrulation. Beginning 4 hpf at the sphere stage of development, it takes zebrafish embryos approximately 6 hours to progress through gastrulation into neurulation (Kimmel et al. 1995).

The main function of gastrulation is to establish the three germ layers (ectoderm, endoderm and mesoderm) and through cell migration rearrange them to establish the main body plan of the embryo (reviewed in Lepage & Bruce 2010). There are four main movements involved in zebrafish gastrulation. The first to begin is epiboly, as the embryo proliferates the cells migrate towards the vegetal pole, eventually completely covering the yolk cell. By 6 hpf the other three main movements have begun. The lateral cells begin to migrate dorsally towards the midline of the embryo in a process called convergence, narrowing the embryo dorsoventrally but extending its anterior posterior axis. The lip cells begin to migrate upwards along the yolk cell of the embryo in a process called involution. This allows the mesodermal and endodermal cells to migrate along the entire axis of the zebrafish embryo. Lastly dorsal cells migrate upwards in extension, this movement extends the anterior posterior axis (reviewed in Solnica-Krezel & Sepich 2012).

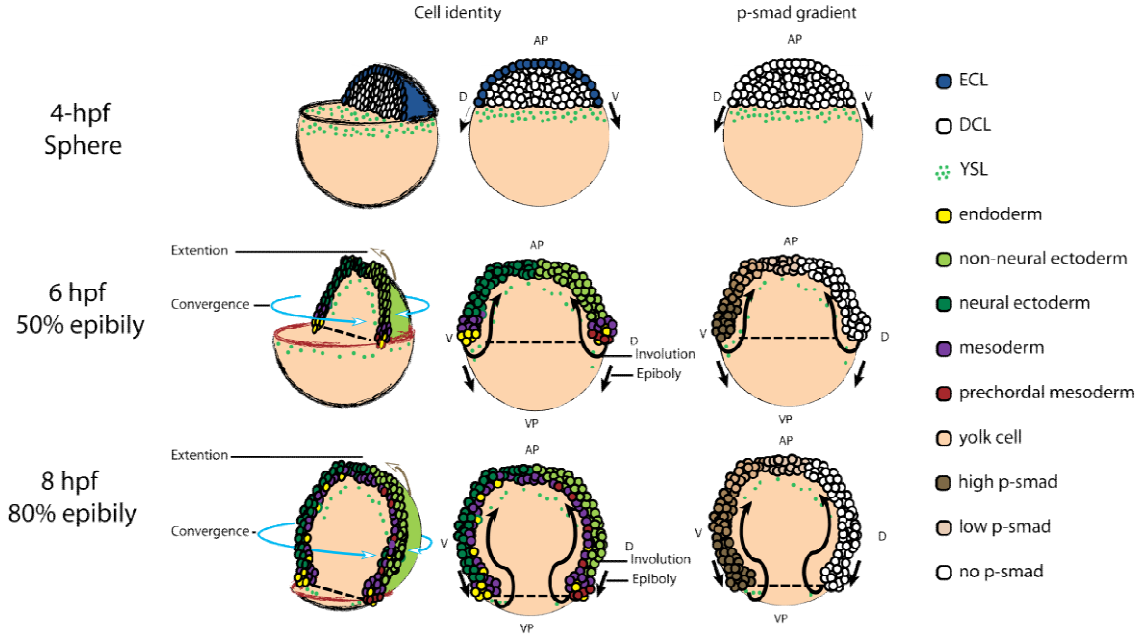



**Fig 3.1.2 zebrafish gastrulation**

The diagram depicts both cell identity of gastrulating zebrafish embryos and the p-smad gradients that specifies the ventral fate of zebrafish embryos.

Figure constructed from information in Solmica-Krezel 2006 and whole mount in situ **Fig 3.2.16.**

**V = ventral D= Dorsal AP= animal pole VP= Vegital pole DCL= Deep cell layer ECL= Enveloping cell layer YSL=yolk syncytial layer**

---

During this process bmp is known to perform two roles. The first identified through mutagenic screens was its role in specifying ventral identity in the dorsal ventral axis. Several bmp pathway mutants were identified which showed varying degrees of dorsalisation, typically an expansion of dorsal / anterior structures (e.g. an expanded notochord) at the expense of the ventral /posterior structures (e.g. a reduced postnatal anal tail) (Mullins et al. 1996). The most severe phenotype observed were the *swirl* and *snailhouse* mutants, later identified as mutations in *bmp2b* and *bmp7* respectively.

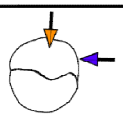
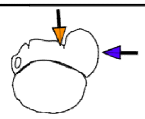
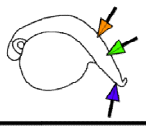
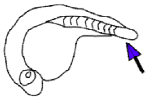
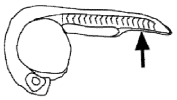
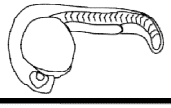
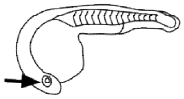
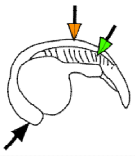
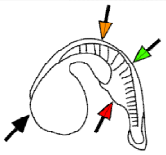
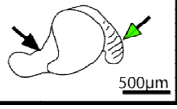
| Scale | Diagram | BMP levels | Chordino levels | Discription |
|---|---|---|---|---|
| C5 |  | ↓ | ↑ | Heavily dorsalised -<br>Enlarged nortochord (↑)<br>The post anal tail is absenf (↑) |
| C4 |  | ↓ | ↑ | Heavily dorsalised -<br>Enlarged nortochord (↑)<br>The post anal tail is absenf (↑) |
| C3 |  | ↓ | ↑ | Dorsalised -<br>Enlarged nortochord (↑) Loss of somites (↑) The tail is reduced and has a charateristic kink (↑) |
| C2 |  | ↓ | ↑ | Dorsalised -<br>The post anal tail is reduced (↑) |
| C1 |  | ↓ | ↑ | Mildly dorsalised -<br>No ventral tail fin (↑) |
| WT |  | | | |
| V1 |  | ↑ | ↓ | Mildly ventralised -<br>Minor neural defects. Most typically reduced eye size (↑) |
| V2 |  | ↑ | ↓ | ventralised embryo -<br>Reduced or absent eyes (↑) and nordochord (↑)<br>Expanded postirour somites (↑) |
| V3 |  | ↑ | ↓ | Heavyly ventralised -<br>No head (↑) No nordochord (↑)<br>Expanded postirour somites (↑)<br>Expanded blood islets (↑) |
| V4 |  | ↑ | ↓ | Heavyly ventralised -<br>No anterior structure (↑)<br>Visble somites (↑) |

**Fig 3.1.3 The Kishimoto scale of zebrafish dorsalisation**
The diagram showing both the key morphological and morphogen differences detailed in Kishimoto et al. 1997.

It was initially assumed that the bmp pathway mutants identified in the screens were representative of incremental changes in the level of bmp activity

(Mullins et al. 1996). This was confirmed when Kishimoto et al. (1997) demonstrated that not only could *bmp2b* nulls be rescued with *bmp2b* mRNA, each of the less severe dorsalisation phenotypes could be recapitulated with different dosages of *bmp2b* mRNA (*swirl* mutants ; Kishimoto et al. 1997). At high doses *swirl* embryos could be made to exhibit ventralised phenotypes, even emulating the complete ventralisation observed in *Chordin* null embryos. From this data they proposed a scale of phenotypes which would encapsulate the various levels of smad activation **(Fig. 3.1.2;** Kishimoto et al. 1997). This scale was used to successfully identify many genes involved in regulating the dorsal ventral fate via the bmp pathway. Examples include the *smad5* (Hild et al. 1999) , *Bambi* (Delot et al. 1999) and *Alk8* (Bauer et al. 2001).

Later work confirmed the existence of a p-smad gradient originating from the dorsal side of the gastrulating embryo. This gradient is primarily mediated by *bmp2b*, as bmp2b initiates the expression of both *bmp4* and *bmp7* in the gastrulating embryo (Ramel & Hill 2012; Kondo 2007). If *bmp2b* is knocked out, no smad phosphorylation occurs (von der Hardt et al. 2007).

To specify the ventral identity bmp signalling must down regulate dorsal genes and upregulate the ventral genes. One critical downstream target is *eve1*. It is initially expressed in response to bmp signalling during the late blastula stage in the ventral mesodermal progenitor cells. Later it is expressed in the somites of the post anal tail (Pyati et al. 2005) and this expression is dependent upon continued bmp expression (Pyati et al. 2005). A zebrafish ventral homeobox transcription factor, eve1 specifies ventral identity, which has been demonstrated by *eve1* overexpression. This results in ventralised embryos (Seebald & Szeto 2011; Barro

et al. 1995), while those that are injected with *eve1* targeting morpholinos are dorsalised, though only if co-injected with morpholinos targeting *vent* , *vox* or *ved*, as *eve1* acts cooperatively with these factors (Seebald & Szeto 2011).

Another well characterised ventral target gene is *gata2*. Originally identified in *Xenopus* as a downstream target of xbmp-4 (Maeno et al. 1996), G*ata2* has since been shown to be a key player in establishing ventral identity in *Xenopus*. Inhibiting *gata2* activity by expressing a dominant negative gata2 protein results in severely dorsalised *Xenopus* embryos(Sykes et al. 1998). In zebrafish it is expressed ventrally by segmentation (10-13 somites; Vuilleumier et al. 2010) and while the precise modulators of expression remain unknown, it is inhibited by I-smad6, demonstrating that it is the downstream target of the canonical bmp cascade (Oren et al. 2005).

Two examples of dorsally expressed genes that are down regulated by bmp signalling are *otx2* and *runx3* (runt like transcription factor). runx3 is known to be a downstream component of tgf-beta and bmp signalling where it acts as a cofactor for the smads. In zebrafish development it is expressed in the early embryo at the mid blastula stage, but it is heavily up-regulated with the onset of neurulation (Cohen 2009). In situ hybridisation experiments have shown that its expression is confined to anterior neural tissue (the trigeminal ganglions, the dorsal neural tube, Rohon–Beard sensory neurons and neural crest cells; Kataoka et al. 2000). Which makes it an ideal marker of dorsally-anterior identity.

*Otx2* is a homeobox transcription factor that is involved in specifying both the development of the vertebrate brain and the initial establishment of the eye field. The function of *otx2* is well conserved throughout vertebrate evolution. It has

been shown to be crucial for normal eye development in *Xenopus* (Zuber et al. 2003), humans (Ragge et al. 2005) and mice (Martinez-Morales et al. 2001). In zebrafish *otx2* has a closely related paralogue *otx1a* which has a similar expression pattern. This accounts for the fact that knocking down *otx2* with MOs results in only minor ocular defects. However double knockdowns (both with MOs and genetic knockouts) show both coloboma and undifferentiated RPE (Lane & Lister 2012; Bharathan, 2014). In zebrafish, *otx2* is under repression from bmp signalling. This is demonstrated by Nikaido et al. who showed that beads coated in *bmp4* suppressed *otx2* expression. This was further confirmed by the fact that cells expressing constitutively active *bmpr1a* (CA-BRIA) had lower levels of *otx2* expression (Nikaido et al. 1999).

The other known function of the p-smad gradient formed during gastrulation is to control the convergence movement of the lateral cells during epiboly. The lateral cells migrate down the bmp gradient towards the dorsal side of the embryo (von der Hardt et al. 2007). This is a completely separate function to the dorsal ventral axis specificity as it doesn't affect the expression of either dorsal or ventral markers. However the phenotype is remarkably similar to the C2 and C3 bmp pathway mutants (Topczewski et al. 2001). Loss of lateral convergence leads to a reduction in the anterior posterior axis manifesting in a reduced post anal tail (Sepich et al. 2000). In fact the *knypek* mutant was initially identified as a bmp interactor because it displays a classic C2 phenotype. Although further investigation showed lateral cell fate was unaffected, and transplanted cells differentiated correctly despite the fact that they had not migrated to the correct

spatial position in the embryo (Topczewski et al. 2001). This similarity is due to the fact that convergence is disrupted in classical bmp mutations and is a significant contributor to the phenotype of mildly dorsalised embryos.

bmp signalling is also a key component of zebrafish eye development. Initially bmp is involved in the establishment of the eye field by antagonising anteriorly expressed eye genes like *otx2* and *rx3* (Bielen & Houart 2012) which is why embryos that have been ventralised have reduced eyes or the eye field is completely absent (Schulte-Merker et al. 1997; Kishimoto et al. 1997; Branam et al. 2010). However, once the eyefield has been established, a new bmp gradient begins to form within the RPE. This pattern is clearly bmp dependent and treatment with the improved bmp inhibitor LDN 193189 leads to ablation of *tbx5* (which is the earliest marker of the dorsal cell fate) expression, but not in *rx3* (the earliest marker of the optic cup) expression. *Bmp2b, bmp4* and *gdf-6* have all been shown to be expressed within the zebrafish RPE (Kruse-Bend et al. 2012). Additionally, up regulating bmp signalling by expressing a constitutively active type 1 bmp receptor in the eye (*rx3:caBMPR1a*) suppresses the expression of the ventral marker gene *vax2,* while suppressing bmp expression in the eye by expressing i-smad7 (*rx3:smad7*). This inhibits the dorsally expressed genes (for example *tbx5* and *Bambi* ;French et al. 2009) but it does not lead to expansion of ventrally expressed genes (for example *vax2* ;French et al. 2009). Both *bmp2b* and *gdf-6* knockdowns have been shown to have independent ventralising activity. When either is knockdown using MOs, the zebrafish retina loses its dorsal identity. This was demonstrated by the expansion of the ventral retinal marker *vax2* and a

loss of the dorsal retinal markers *tbx5*, *tbx2b*, and *efnb2* (Gosse & Baier 2009; French et al. 2009; Kruse-Bend et al. 2012). *W*hile *bmp4* depletion does not have an observable effect on the zebrafish retina, overexpression of *bmp4* does lead to an expansion in *tbx5* expression, but only in the presence of gdf-6(Kruse-Bend et al. 2012).

## 3.1.5 smoc-1 is linked BMP signalling

In *Xenopus*, xsmoc-1 has been shown to act as an antagonist of BMP signalling. Both *xsmoc1* depletion by morpholino oligonucleotide injection and overexpression by *xsmoc1* mRNA injection show the characteristic phenotype of a bmp antagonist. Even more compellingly, *xsmoc1* overexpression can be rescued if the embryos are injected with a constitutively active bmp receptor. When *xsmoc1* is only depleted unilaterally, the *Xenopus* develops unilateral anophthalmia on the side of the injection (Thomas et al. 2009).

The most well studied model of bmp signalling gradients is the *Drosophila* wing imaginal disc. Decapentaplegic (Dpp) is the *Drosophila* homologue of the Bmps. In this system the *Drosophila* homologue of smoc-1, Pentagone (Pent), was identified bioinformatically as a sequence suppressed by Dpp. Pentagone acts as an antagonist of Dpp signalling by controlling the rate of Dpp endocytosis (Vuilleumier et al. 2010). The abnormal gradient in *Pentagone* mutants can be rescued by knocking down the Dpp receptor *Thickveins*. This suggests that the

gradient pattern formation is endocytosis-dependent. If this were not the case, the mad signal would get weaker but the shape of the gradient would remain unchanged. A study by Vuilleumier *et al.* indicates that Pentagone acts through an interaction with Dally (Vuilleumier et al. 2010), and as has previously been discussed in **3.1.3** Dally is already known to regulate endocytosis of bmp receptor complexes (Akiyama et al. 2008).

Vuilleumier *et al.* showed that Pentagone expression was down regulated by Dpp. This mutual antagonism suggested that they operated in a feedback loop, each suppressing the activity of the other. Later this was confirmed by Hamaratoglu et al. (2011). In wild type *Drosophila* imaginal disc, the p-Mad gradient will remain constant with the growth of the disc. This is called scaling. However, in *Pentagone* mutants the width of the p-smad gradient remains unchanged even as the wing disk triples in size over the course of its development (Hamaratoglu et al. 2011) which demonstrates the function of the Pentagone/Dpp feedback loop is to control the shape of the p-MAD gradient. It would allow the Dpp gradients to scale with the growing organism and make the gradients much less susceptible to alteration in external conditions than wild type flies (Hamaratoglu et al. 2011).

A defect in BMP signalling is a plausible for OAS in humans. BMP has been long been known to be a central player in vertebrate eye development, and mutations in *BMP4* have been shown to cause both anophthalmia and limb defects in humans (Bakrania et al. 2008).

## 3.1.6 The current evidence for a link between smoc1 and bmp signalling in zebrafish

Published work on *zsmoc1* is very limited. In situ hybridisation has shown that smoc1 expressed in the ventral RPE of the developing eye (Abouzeid et al. 2011). Abouzied et al. also reported that MO depletion of *zsmoc1* lead to an increased incidence in coloboma, both of which are consistent with a function regulating the bmp gradient which establishes dorsal retinal identity in the zebrafish eye.

Pilot work conducted during the Masters project preceding this thesis strongly implicated *zsmoc1* function in bmp signalling (David Sexton 2010). Initial work showed that *zsmoc1* morphants had the classical moderately dorsalised phenotypes **(Fig. 3.1.3 A, B, C and D; Fig. 3.1.2).** A link with the bmp pathway was confirmed by coinjection with *bmpr1a* which rescued the dorsalisation of *zsmoc1* morphants **(Fig. 3.1.3 E),** and that this involved a canonical BMP signalling as whole mount immunohistochemistry showed p-smad phosphorylation was dramatically upregulated in *zsmoc1* morphants.

The bmpr1a rescue strongly suggests the morphant phenotype is not due to off-targeted effects. Yet, no rescue was achieved using *zsmoc1* mRNA during the project. Without an mRNA rescue off-target effects cannot be ruled out conclusively (Skromne & Prince 2008).

---

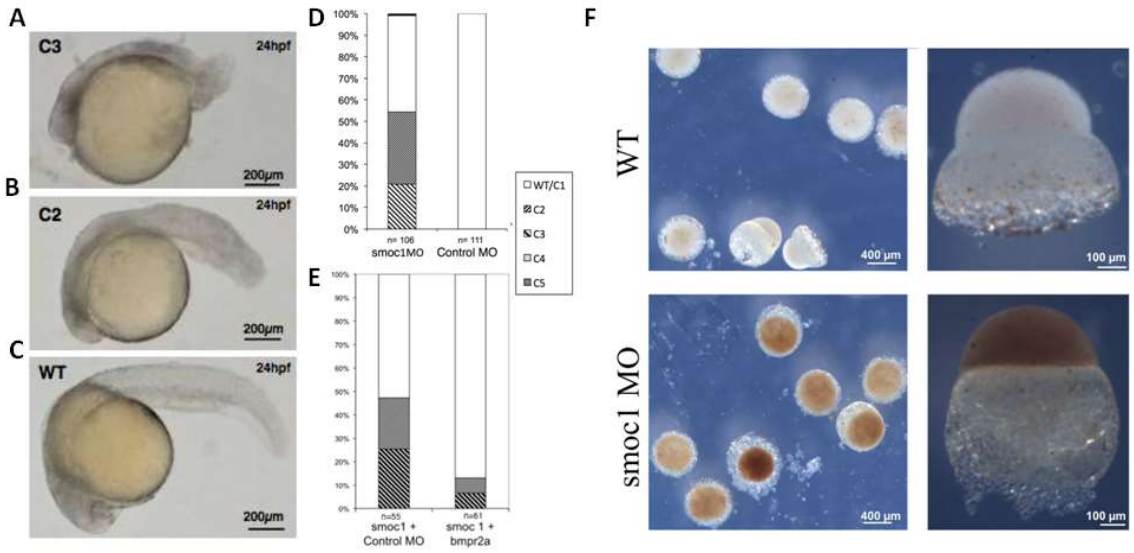



**Fig 3.1.4 Anteroposterior and dorsoventral patterning in zebrafish embryos requires z*smoc1*.**

One day after smoc1aMO injection embryos show bmp dorsalisation phenotype **(A)** A 1 day old C3 dorsalised embryo showing embryonic length, almost complete loss of tail and bent notochord and visible neural structures. **(B)** A C2 dorsalised embryo with a less severe reduction to be post anal tail **(C)** A phenotypically normal fish. **(D)** the smocMO1a injected embryos showed disrupted dorsoventral patterning while control injected embryos were phenotypically normal. smoc-1 depletion up regulates p-smad signalling. smoc1 embryos stained immunohistochemistically shows p-smad is upregulated in smoc1MOa injected embryos after 4 hpf. **(H)** bmpr2aMO coinjected with zsmoc1aMO can rescue the dorsalisation of *zsmoc1* morphants. **(F)** smoc1 embryos stained immunohistochemistically shows p-smad is upregulated in smoc1MOa injected embryos after 4 hpf.

(David Sexton 2010)

| Name | 1. Type | 2. Sequence | 3. Knockdown | 4. Used in |
|---|---|---|---|---|
| zsmoc1saMO | Splicing | CACATAAATGAAGCGATGAGGGCAT | Yes (Sexton 2010) | (Sexton 2010) |
| zsmoc1sbMO | Splicing | CACATAAATGAAGCGATGAGGGCAT | *Yes | (Sexton 2010) Fig 3.2.16 |
| zsmoc1scMO | Splicing | ATGCATACAGTACCTGAACACACT | n/a | (Sexton 2010) Fig 3.2.1, Fig 3.2.4. |
| zsmoc1tMO | Translation | AAGAGCCAGATTGTGACAGTTCATC | n/a | Fig 3.2.1, Fig 3.2.4, Fig 3.2.7, 3.2.14 and 3.2.15. |

| Name | 5. p53 MO coinjection | 6. Rescued |
|---|---|---|
| zsmoc1saMO | no | no |
| zsmoc1sbMO | no | no |
| zsmoc1scMO | yes - Fig 3.2.1, Fig 3.2.4. | no |
| zsmoc1tMO | yes - Fig 3.2.1, Fig 3.2.4, Fig 3.2.7, 3.2.14 and 3.2.15. | yes (Fig 3.2.7) |

**Table 3.1.3 Summary on the zsmoc1 targeted morpholinos used during project**

Table summarises key properties morpholinos and details of the experiments they were used in. **1.** Shows whether the MO is blocks splicing or translation, **2.** Shows the Morpholinos base sequence, **3.** Details whether morpholinos has been shown to reduce the levels of correctly spliced *zsmoc1* mRNA experimentally, **4.** Shows which in figures a specific MO was used, **5.** Summarises which figures derived their results from embryos coinjected with morpholinos targeting *zsmoc1* and *p53*, **6.** Specifies whether or not a MO has been rescued with hSMOC1 mRNA.

* Result was obtained by Kishan Aldridge using qPCR at 4 hpf embryos

# 3.2 Results

## 3.2.1 Characterising zebrafish *smoc1* (*zsmoc1*) morphant phenotype

To validate the MOs I had designed to target *zsmoc1*, it was important to compare the induced phenotype to that of previous reports in the literature. This was particularly relevant for the translation blocking MO that I had designed to the newly identified exon 1 **(chapter 2.2.5).** Both of my *zsmoc1* targeting MOs (one blocking translation and the other splicing) were individually co-injected with a *p53* blocking MO to rule out off target effects caused by the triggering of the *p53* response. Even with the co-injection of *p53,* the same characteristic spectrum of bmp phenotypes was observed. Indeed the new translation blocking MO (zsmoc1tMO) are more dorsalised than those injected with the splice blocking MO (zsmoc1cMO). Only 12% of zsmoc1tMO morphants were phenotypically normal compared to 43% zsmoc1cMO **(Fig. 3.2.1 A; B).** Additionally 18% of embryos injected with the smoc1tMO showed the more severe C4 phenotype.

To further establish if the cells were indeed dorsalised, a quantitative RTPCR assay (qPCR) was established using the relative expression of bmp downstream targets. For validation purposes it was necessary to produce positive control morphant embryos treated with known bmp agonists and antagonists.

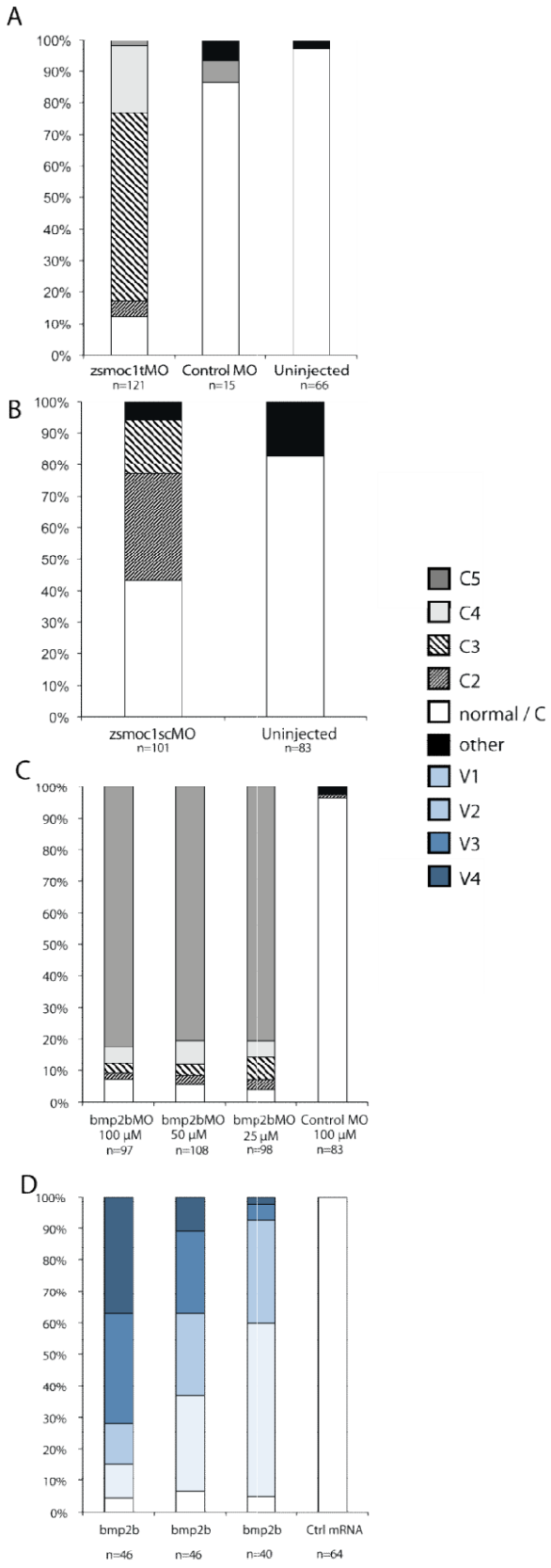



**Fig. 3.2.1 New *zsmoc1* targeting morpholinos induce the characteristic dorsalisation phenotypes previously reported in the media. Even when co-injected with *p53* targeting MO**

**(A)** The new trans targeting MO zsmoc1tMO shows the same range of dorsalisation previously reported. **(B)** Injection of splice morpholino zsmoc1c-MO recapitulates the spectrum of phenotypes previously reported even with the coinjection of p53 MO and mirrors the defects seen in the trans MO.

*bmp2b* morphants show the characteristically severe (C5) dorsalisation phenotype previously reported in Lele et al. (2001). Confirming they are biologically active **(C)**.
Injection of *bmp2b* mRNA also showed the characteristic ventralised phenotype previously reported in Kishimoto et al. (1997) **(D)**.

The previously published bmp2b MO was chosen for use as a positive control for dorsalisation (Lele et al. 2001). In our hands it recapitulated the severely dorsalised phenotype previously reported at all dosages tried **(Fig. 3.2.1 C).** Confirming it is a suitable control for the qPCR assay. *bmp2b* mRNA was selected as the positive control for ventralisation as it had previously been shown to induce highly ventralised embryos (Kishimoto et al. 1997). The mRNA showed a strong dose-response, with higher concentrations of mRNA leading to more strongly ventralised phenotypes (V3,V4) **(Fig. 3.2.1 D)**. Confirming *bmp2b* mRNA is a valid positive control for ventralisation.

## 3.2.2 Developing a quantitative RTPCR assay to assess dorsalisation and ventralisation in morphant zebrafish embryos

To assess dorsalisation and ventralisation, we selected several downstream targets of bmp. Two sets of genes were chosen. One of neural markers comprised of both *otx2* and *runx3* which are normally expressed in the dorsal region of the early embryo (Navaneetha Krishnan Bharathan, 2014; Kataoka et al. 2000; Vuilleumier et al. 2010). The other set of genes are markers of normal ventral ectoderm. *eve1* and *gata2* expressed are in the ventral ectoderm of the neurulating zebrafish embryo (Seebald & Szeto 2011, p.1; Oren et al. 2005, p.2; Vuilleumier et al. 2010).

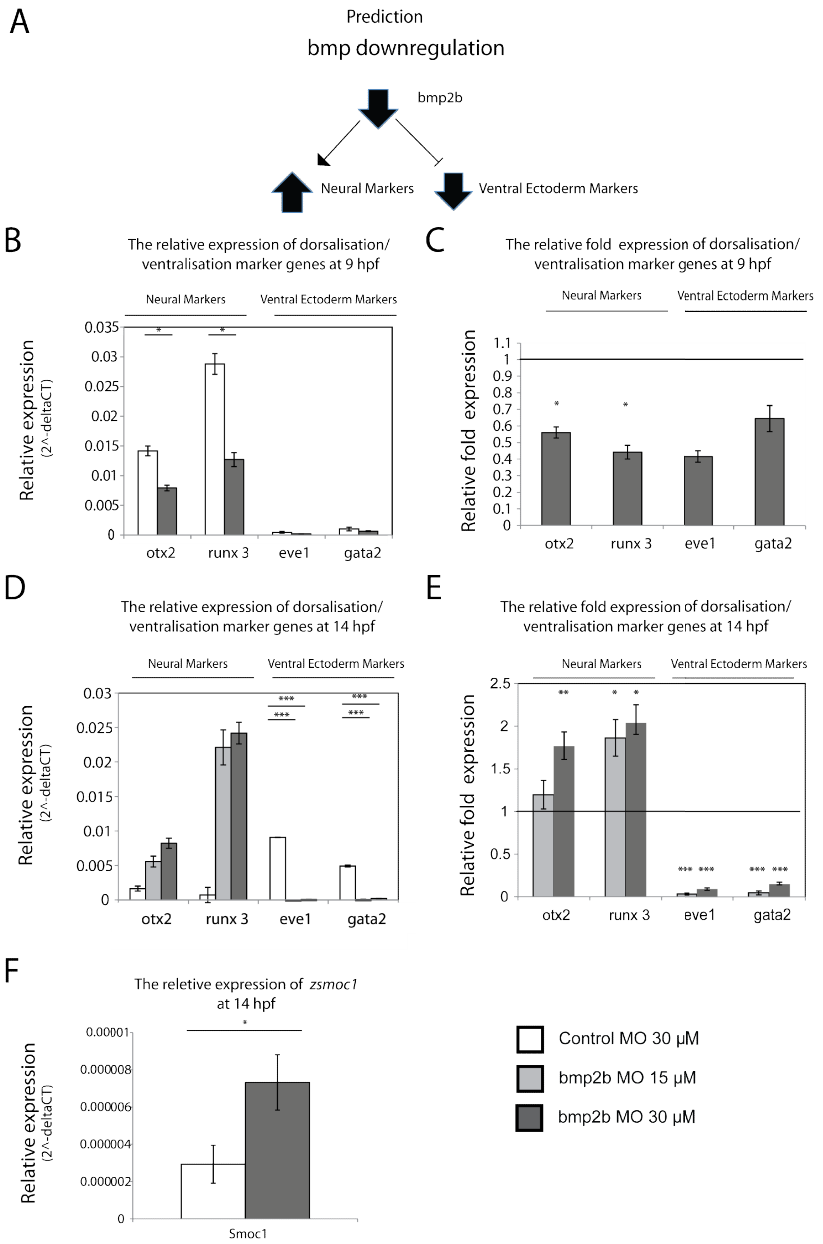



**Fig. 3.2.2 Down-regulation of bmp signalling in *bmp2b* morphants leads to the predicted reduction in the expression of ventral ectoderm markers and a corresponding up-regulation of *smoc1* and neural markers at 14hpf but not at 9hpf.**

**(A)** Schematic representation of the predicted response of neural and ventral ectoderm genes when bmp signalling has been down regulated (i.e. by the injection of bmp2b MO) in zebrafish embryos. *bmp2b* morphants do not show the predicted pattern of neural markers and ventral ectoderm expression predicted for an ablation of bmp signalling at 9 hpf **(B; C).** However , at 14 hpf the expression pattern is entirely consistent with the one predicted **(D; E).** The expression of neural markers is increased while the expression of ventral ectoderm markers is reduced. **(F)** Additionally *zsmoc1* is up regulated at 14 hpf in *bmp2b* MO.

**(B; D; F)** Gene expression levels were calculated using the comparative Ct method and normalised to *efla.* **(C; E)** Is relative fold expression in relation to zebra fish injected with 30 µM Control MO. All samples were analysed in triplicate. (mean ± s.e.m) and all p-values were calculated using a type 2 parametric students ttest * p<0.05 ** p<0.001 *** p<0.001

If the C and V phenotypes are indeed indicative of dorsalisation then we predicted that an up regulation of bmp signalling should result in a down regulation of the neural marker genes and a corresponding up regulation of the ventral ectoderm genes **(Fig. 3.2.1 A)**.To test this, bmp2b signalling was down regulated in zebrafish embryos by injecting with bmp2b MO. The embryos were collected at 9hpf and 14hpf. The mRNA was extracted and transcribed to cDNA. The releative levels of expression of the marker genes was quantified with qPCR. At 9 hpf both ventral and neural markers were down regulated **(Fig. 3.2.2 B; C)**. However, by 14hpf the predicted pattern of marker gene expression was observed **(Fig. 3.2.2 A; D; E)**. The dorasally expressed neural markers were upregulated while the ventrally expressed ectoderm markers were downregulated. This confirms that *bmp2b* morphants are actually dorsalised. As the markers only followed the predicted pattern at 14 hpf, it was decided that for future dorsoventralisation assays every qPCRs would be conducted on this time point.

To confirm that the marker genes also showed the predicted expression patterns for ventralised embryos the reciprical experiment was performed. *bmp2b* expression was upregulated through the injection of *bmp2b* mRNA into zebrafish embryos **(Fig 3.2.3. B; C)**. This led to a down reglation of *otx2* and *runx 3* and upregulation of *eve1* and *gata2,* which is the expected gene expression pattern of ventralised embryos, confirming that these markers can distinguish between both dorsalised and ventralised embryos.

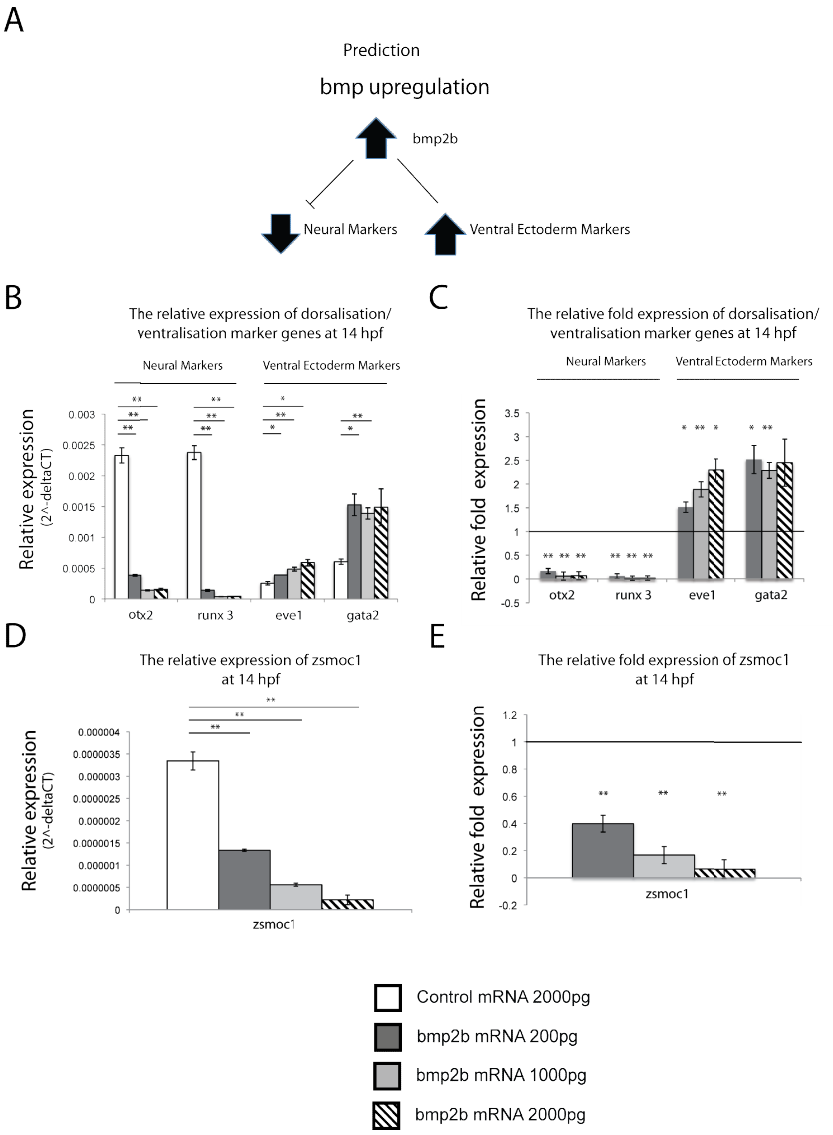



**Fig. 3.2.3 Upregulating bmp signalling by overexpressing *bmp2b* mRNA leads to an upregulation of ventral ectoderm markers and a corresponding down-regulation of *zsmoc1* and neural markers at 14hpf.**

**(A)** Schematic representation of the predicted response of neural and ventral ectoderm genes when bmp signalling has been up regulated (i.e. by the injection of *bmp2b* mRNA) in 14 hpf embryos. **(B; C)** Injection of *bmp2b* mRNA leads to a down regulation of neural markers and a corresponding up regulation of ventral ectoderm genes predicted for an up regulation of bmp signaling. **(D; E)** *bmp2b* over expression leads to a corresponding down regulation in zsmoc1 expression in a dosage dependent manner.

**(B; D)** Gene expression levels were calculated using the comparative Ct method and normalised to *elfa*. **(C: E)** Is the relative fold expression in relation to zebra fish injected with 2000pg of antisense *bmp2b* mRNA.
Samples were analysed in triplicate (mean ± s.e.m) and all p-values were calculated using a type 2 parametric students ttest * p<0.05 ** p<0.001 *** p<0.001

Additionally *zsmoc1* expression is dependent on to the level of bmp activity. In *bmp2b* morphants *zsmoc1* expression is upregulated **(Fig 3.2.2 F)**, conversly the injection of *bmp2b* mRNA leads to a dose dependant reduction of *zsmoc1* levels **(Fig. 3.2.3 D; E)**. This expression pattern is strongly indicatve of a feedback loop (Vuilleumier et al. 2010).

## 3.2.3 The effect of zsmoc-1 depletion on ventral ectoderm markers and dorsal neural markers

Although *zsmoc1* morphants were scored as having the C2 –C4 phenotypes, they do not show any increased expression of the dorsalisation markers. They also failed to show a typical expression pattern of a bmp antagonist, as zsmoc1tMO showed a universal reduction in the expression of both dorsalisation and ventralisation markers **(Fig. 3.2.4 A; B)**. This indicates that the *zsmoc1* phenotype is not the result of "typical" dorsalisation. To rule out off target effects, the results were replicated with the splice targeting MO zsmoc1scMO **(Fig. 3.2.4 C; D)**, although as with the phenotypic analysis **(Fig. 3.2.1 A; B),** zsmoc1tMO had a more severe effect.

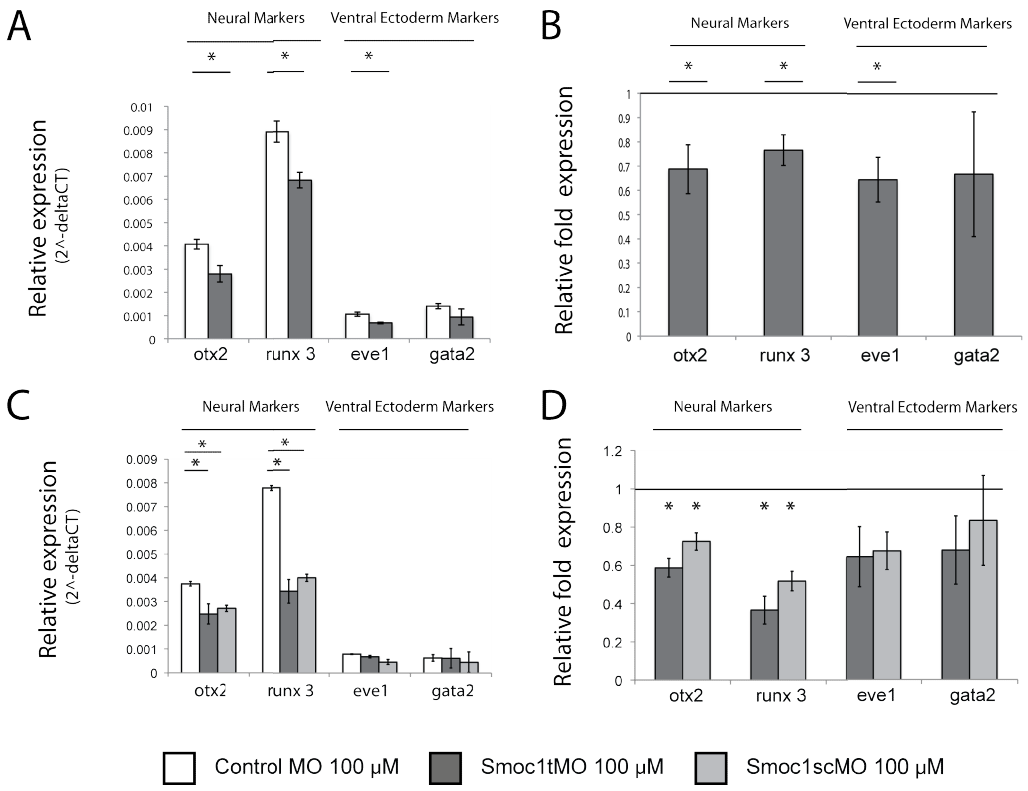



**Fig 3.2.4 The neural and ventral ectoderm marker genes are universally down regulated in 14 hpf *zsmoc1* morphants.**

**(A)** Embryos injected with the translation blocking zsmoc1 targeting MO zsmoc1tMO showed a reduced levels of both neural and ventral ectoderm markers. This change was recapitulated in both *otx2* and *runx3* for embryos injected with smoc1scMO **(C; D)**. Though *eve1* and *gata2* were both down regulated in embryos injected with either smoc1tMo and smoc1scMO these differences are not statistically significant.

**(A; B)** Results are the mean (± s.e.m) of three biological replicates carried out in triplicate. p-values were calculated using a type 2 parametric students ttest * $p<0.05$ ** $p<0.001$ *** $p<0.001$
**(C; D)** Results are the mean (± s.e.m of three technical replicates. p-values were calculated using a type 2 parametric students ttest * $p<0.05$ ** $p<0.001$ *** $p<0.001$
**(A; C)** Gene expression levels were quantified using the comparative CT method and normalised to *elfa*.
**(B; D)** Show the relative fold levels of expression of A:C relative to embryos injected with 100 µM of Control MO

## 3.2.4 The effect of *zsmoc1* mRNA injection on ventral ectoderm markers and dorsal neural markers

Injection of *zsmoc1* mRNA did not significantly alter the expression of the dorsalisation marker genes **(Fig. 3.2.6).** However, the injection of *hSMOC1 m*RNA did have an effect. *hSMOC1* injection ventralised the zebrafish embryos. The neural marker genes were significantly down regulated, while the ventral ectoderm markers were up regulated **(Fig. 3.2.7 B)**. Which indicates that hSMOC-1 is acting as a BMP agonist. Phenotypically embryos injected with *hSMOC1* mRNA show significantly greater defects than those injected with *bmp2b* RNA. Epiboly is arrested at 10 hpf and the majority of the embryos die before completing neurulation **(Fig. 3.2.7 A)**.

Since the *zsmoc1* mRNA did not appear to be biologically active* we went on to use the *hSMOC1* mRNA to attempt to rescue the *zsmoc1* morphant.

*It should be noted that it later transpired that *zsmoc1* mRNA lacked both exons 14 and 15 which provides a plausible explanation for the absence of a biological effect. Though other explanations cannot be ruled out.

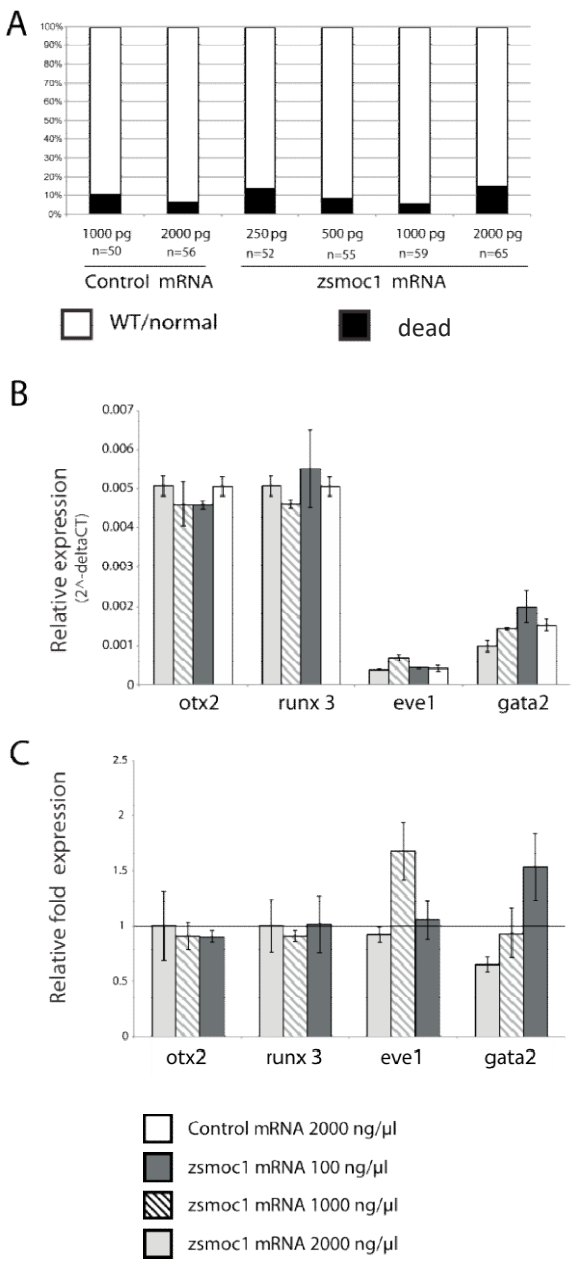



**Fig. 3.2.5 *zsmoc1* mRNA injection does not significantly alter embryo patterning.**

**(A)** Injection of *zsmoc1* mRNA leads to no increase in zebrafish mortality. Embryo survival was scored between 8 hpf (40% epiboly) and 14 hpf (neurulation) **(B)** qPCR was used on cDNA from 14 hpf (10-13 somites) stage embryos showed no statistically significant change in the expression levels of neural and ventral ectoderm markers and ventral ectoderm marker genes. **(C)** qPCR was used to quantify the bmp reporter *id1* and the FGF reporter *dusp6*. All qPCR was normalized to *elfa* and quantified using $2^{-(\Delta\Delta cT)}$. Relative fold expression was compared to embryos injected with an equivalent quantity of control mRNA (1000pg of antisense *hSMOC1*.)

results are derived from 3 technical replicates (mean ± s.e.m)

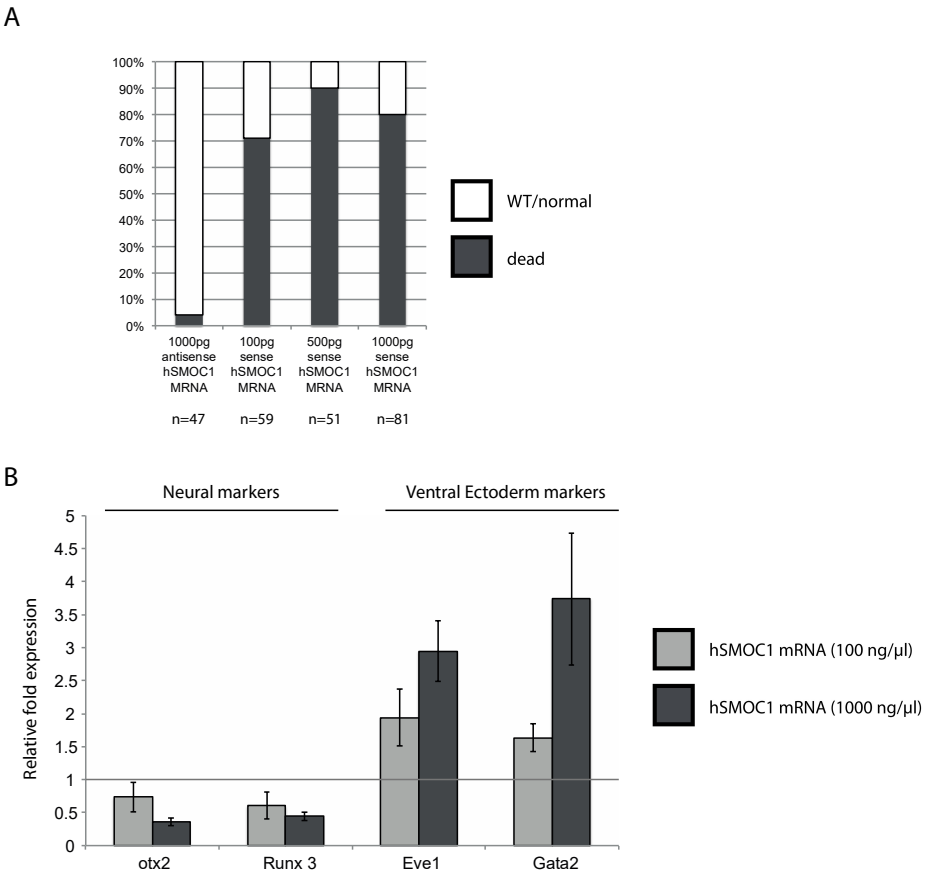



**Fig. 3.2.6 *hSMOC1* mRNA injection leads to defects in embryo patterning consistent with an agonist of bmp signaling.**

**(A)** Injection of over 100pg of *hSMOC1* RNA is lethal to the majority of embryos. Embryo survival was scored between 8 hpf (40% epiboly) and 14 hpf (neurulation) **(B)** qPCR analysis of cDNA from 14 hpf (10-13 somites) stage embryos showed that *hSMOC1* injection leads to a down regulation of *otx2* and *runx3* and an upregulation of *eve1* and *gata2*.

**(B; C)** All samples were measured in triplicate, and the results were quantified using $2^{-(\Delta\Delta cT)}$ and normalized to *elfa*. Relative fold expression was compared to embryos injected with an equivalent quantity of control mRNA (1000 µg/µl of antisense *hSMOC1*.)

## 3.2.5 *hSMOC1* mRNA injection partially rescues zsmoc1MO injection

It remained possible that the universal down-regulation of both the neural and ventral ectoderm markers was attributable to off target effects. To test this hypothesis, *hSMOC1* mRNA was co-injected with zsmoc1tMO to see if this could rescue the down-regulation of the marker genes.

Co-injection of *hSMOC1* mRNA results in partial rescue of *zsmoc1* morphants. Embryos injected with 50pg of *hSMOC1* mRNA show no statistically significant reduction in levels of *otx2*, *eve1* and *gata2* **(Fig. 3.2.7 A and B)**. *Eve1* shows dosage response with 10pg of *hSMOC1* mRNA increasing its expressions towards wild type levels **(Fig. 3.2.7 A and B)** demonstrating that the knockdown of the dorsal-ventral marker genes (*otx2*, *eve1* an *gata2*) caused by zsmoc1tMO injection **(Fig. 3.2.4 A and B)** is due to the knockdown of *zsmoc1* and not off target effects.

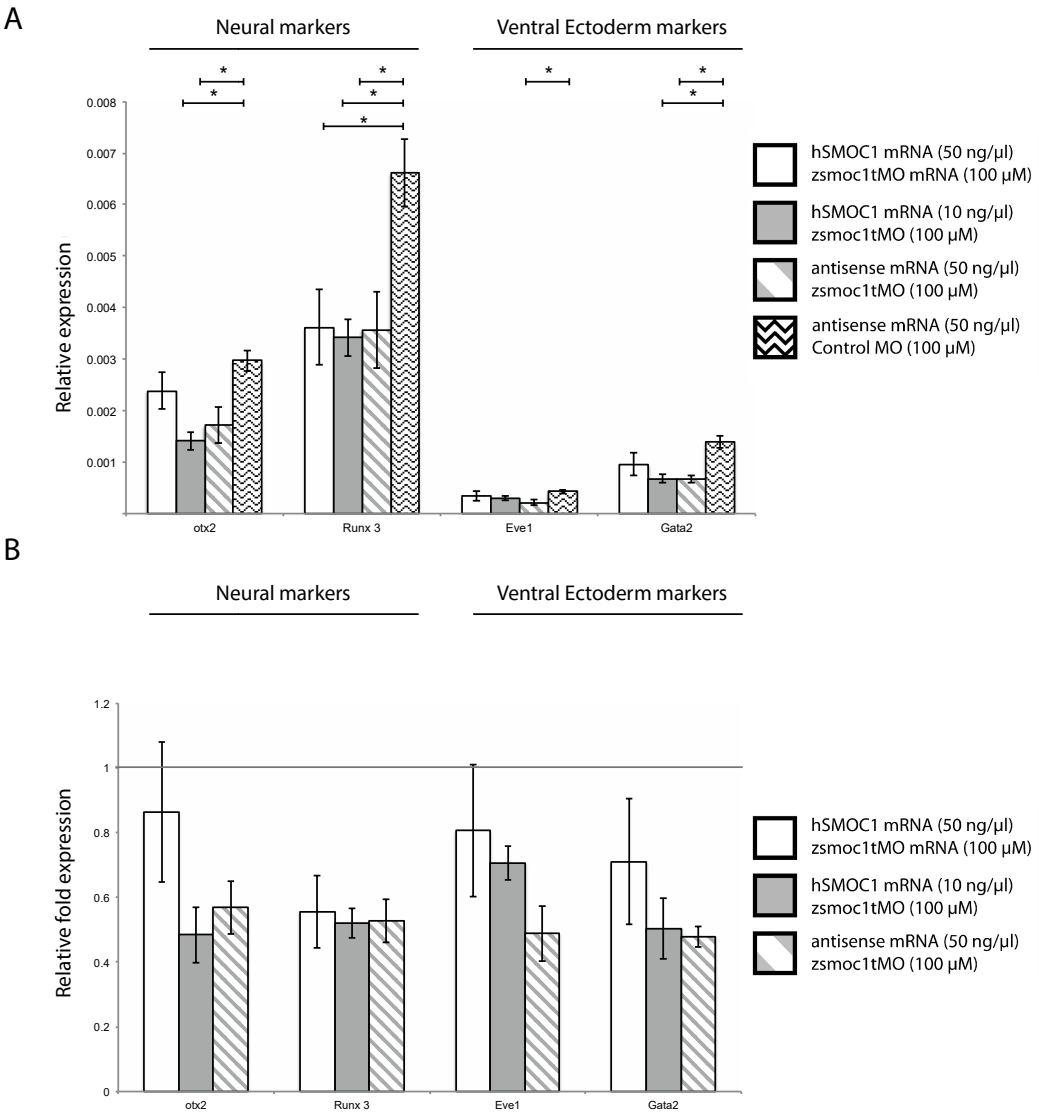



**Fig. 3.2.7 Injection of *hSMOC1* mRNA can partially rescue *zsmoc1* morphants**

Coinjection of 50 pg of *hSMOC1* mRNA results in a partial rescue of the reductions in *otx2 eve1* and *gata2* expression caused by the injection of zsmoc1tMO. *Runx3* levels remained unaffected. **(A)** Relative expression levels ($2^{-(\Delta\Delta cT)}$) normalized to *elfa*. **(B)** Relative fold expression relative to embryos injected with 50 ng/µl antisense *hSMOC1* mRNA and 100 µM of control MO.

Results are the aggregate of 3 biological replicates each with 3 technical replicates (mean =± s.e.m). P values were calculated using 2 tailed type 2 parametric ttests. *=p<0.05 **=p<0.001 ***=p<0.0001

## 3.2.6 Morpholino knockdown of *bmpr2b* gives similar results to morpholino knockdown of *zsmoc1*

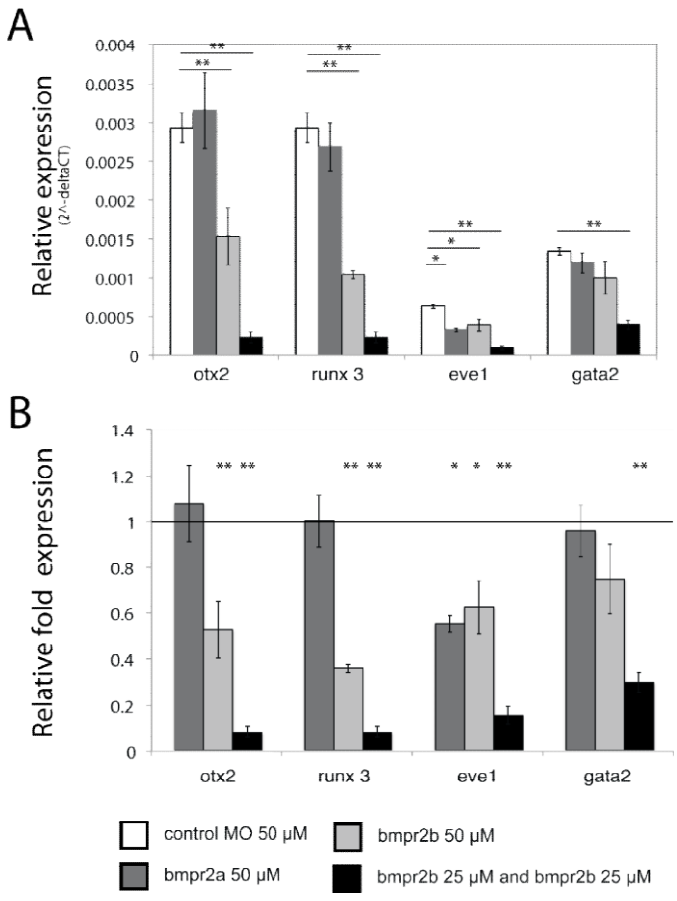



**Fig. 3.2.8 *bmpr2a* and *bmp2a* morpholinos act synergistically to control dorsoventral patterning.**

Injection of bmpr2a MO only down regulates *eve1* while bmpr2b MO results in a significant reduction in *eve1, otx2* and *runx3* expression. When co-injected the down-regulation is enhanced with each marker more severely down-regulated when compared single MO injection. **(A)** The relative expression calculated with the comparative CT method and normalized to *elfa* **(B)** Relative fold expression relative to embryos an equivalent dosage of 50 control MO expression.

Results are the aggregate 3 technical replicates (mean =± s.e.m). P values were calculated using 2 tailed type 2 parametric ttests. *=p<0.05 **=p<0.001

Morpholino knockdown of type2 bmp receptor genes resulted in a similar pattern of dorsal-ventral gene disruption to *zsmoc1* morphants **(Fig. 3.2.4 ; Fig. 3.2.8)**. *bmpr2a* knockdown caused only mild misregulation with only *eve1* being significantly down regulated. While *bmpr2b* knockdown results in reduced *eve1*, *otx2* and *runx3* in a similar patter to that seen in *zsmoc1* morphants **(Fig. 3.2.4 ; Fig. 3.2.8)**. The type 2 bmp receptors display a degree of redundancy in dorsoventral patterning, as shown by coinjection of 25µM of bmpr2a and bmpr2b

morpholinos results in significantly enhanced down regulation of every marker gene.

## 3.2.7 *hSMOC1* mRNA does not rescue either *bmp2b* mRNA induced ventralisation or bmp2b MO induced dorsalisation phenotypes

In an attempt to establish a direct interaction between *hSMOC1* and bmp signalling Co-injections of *hSMOC1* mRNA and either bmp2b MO or *bmp2b* mRNA were performed **(Fig. 3.2.9; Fig. 3.2.10)**. However, neither sets of injections showed rescue. Nor did they display any synergistic enhancement of the expression ventralisation and dorsalisation markers. With coinjection *hSMOC1* mRNA and *bmp2b* mRNA leading to the same degree of ventralisation as is observed in control embryos injected with either *hSMOC1* mRNA or *bmp2b* mRNA alone **(Fig. 3.2.9)**. While coinjection of bmp2b MO with *hSMOC1* mRNA also failed to rescue *bmp2b* morphants **(Fig. 3.2.10)**. These results do not rule out an interaction however as rescue experiments are often difficult to perform (Skromne & Prince 2008).

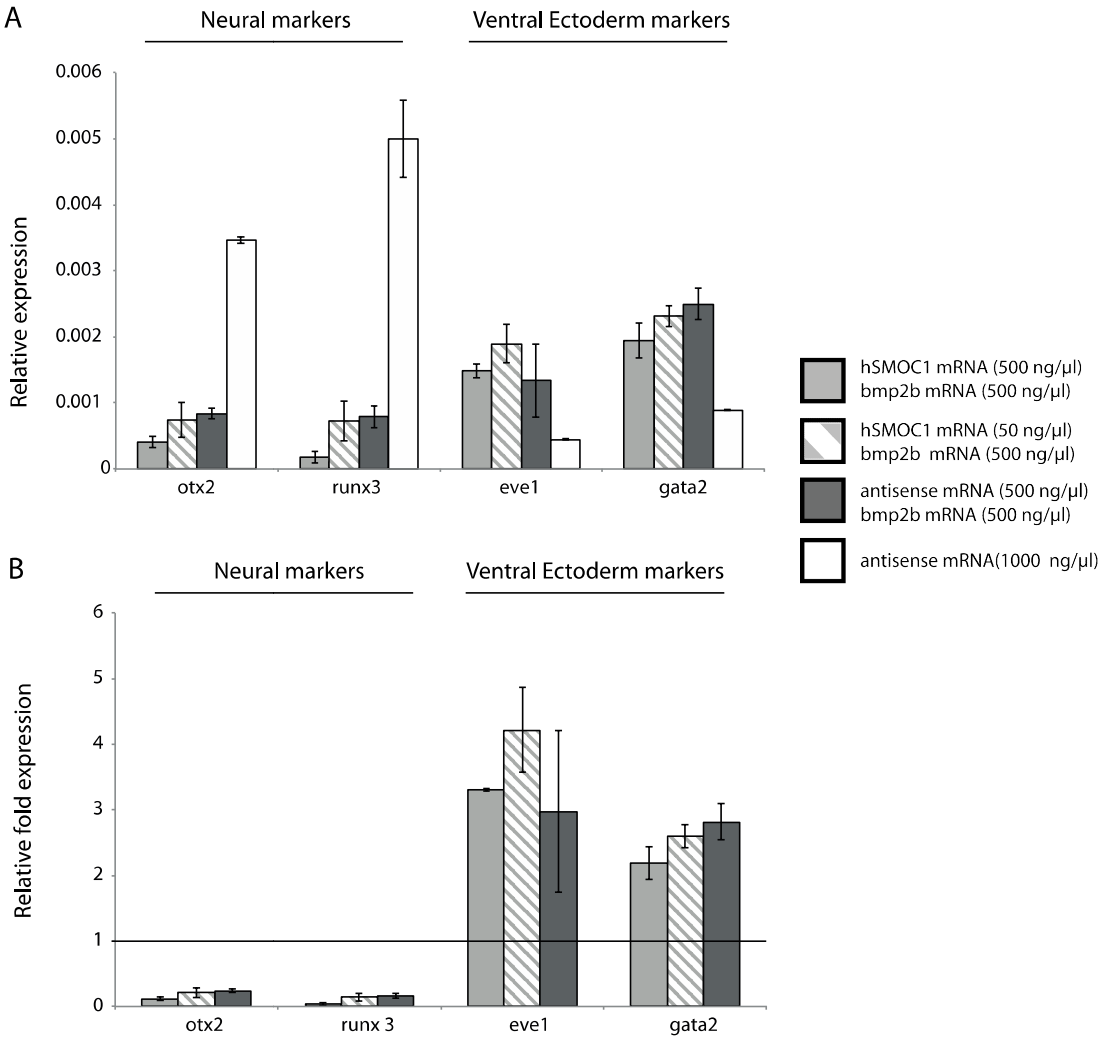



**Fig. 3.2.9 co-injection of *hSMOC1* mRNA with *bmp2b* mRNA does not alter the expression of neural and ventral ectoderm markers**

Co-injection of *bmp2b* mRNA with *hSMOC1* mRNA into zebrafish embryos does not result in significant differences in the relative expression of dorsalisation markers and genes assayed by qPCR. **(A)** Relative expression, the delta Square CT after values have been normalized to *elfa* **(B)** Relative fold expression relative to the Delta square CT embryos injected with 1000 nl/µl of antisense *hSMOC1* RNA. The results are averaged from 2 biological replicates each having 3 technical replicates (mean ± s.e.m).

.

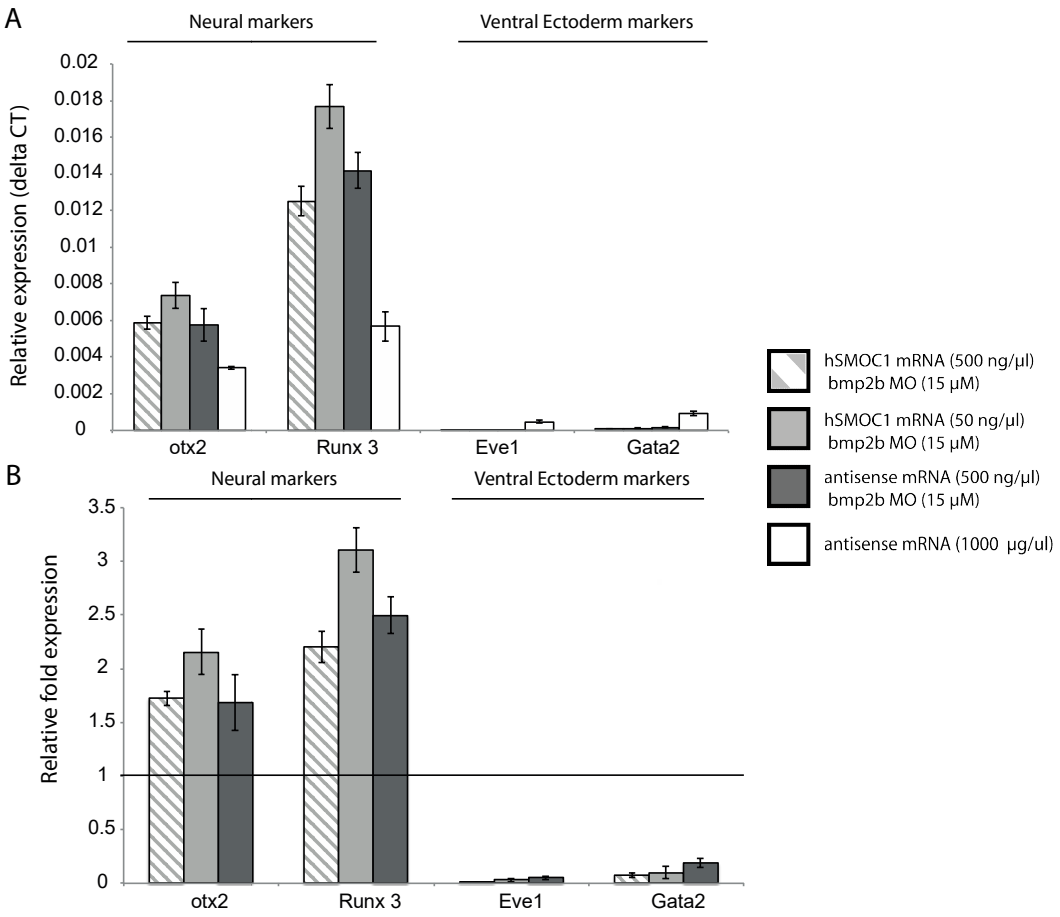



**Fig 3.2.10 The injection of *hSMOC1* mRNA cannot rescue *bmp2b* morphants.**

Co-injection of *hSMOC1* mRNA and bmp2b MO into zebrafish embryos does not result in significant differences in the relative expression of dorsalisation markers and genes assayed by qPCR. **(A)** Relative expression, the delta Square CT after values have been normalized to *elfa* **(B)** Relative fold expression relative to embryos injected with 1000 nl/µl of antisense *hSMOC1* RNA. The results are averaged from 2 biological replicates each having 3 technical replicates(mean ± s.e.m).

# 3.2.7 Optimising automated live imaging protocol for early zebrafish embryos

For the initial attempts at live embryo imaging zebrafish embryos were embedded with their chorions in 2% low melting point agarose when they reached the sphere stage of development (4 hpf **; Fig. 3.2.11)**. Embryos were imaged using brightfield microscopy every 30 min. Under these conditions the embryos develop apparently normally and spontaneously align horizontally. Which would enable the observation of eye development but from only one angle. By 20 hpf the embryos had begun to move, though the movement was not significant.

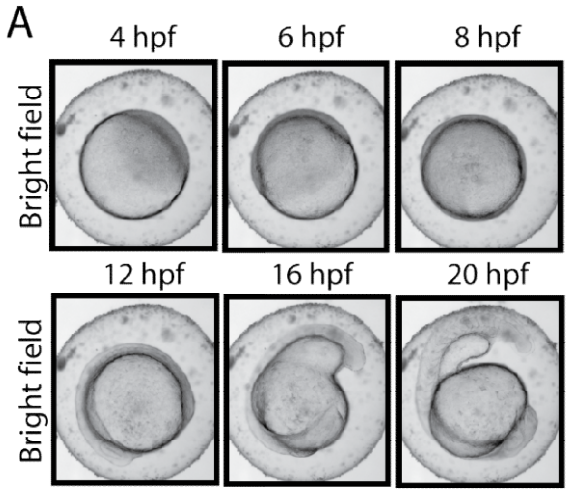



**Fig 3.2.11 Zebrafish embryos embedded in 2% agarose survived overnight incubation if embedded with chorions.**

Embryos were mounted in 2% low melting point agarose and imaged seven times (the plane of focus varied but 30 µm) every 30 minutes. The free movement afforded to them within chorions lead to limited control of embryo orientation. By 8 hpf every embryo lay in the horizontal plane. At 20 hpf embryos started to move within the chorion, but movement was minimal.

Our aim was to record transgenic fluorescent reporter signalling in real-time in the developing eye. To test how the zebrafish embryos responded to UV excitation a transgenic zebrafish line carrying a gfp tagged version of the ubiquitously expressed histone variant *his2avD* (H2A:gfp zebrafish; Pauls et al. 2001) was used. Under UV illumination the embryos were found to move

significantly, rendering accurate observation of the eye field and eye difficult **(Fig. 3.2.12)**.

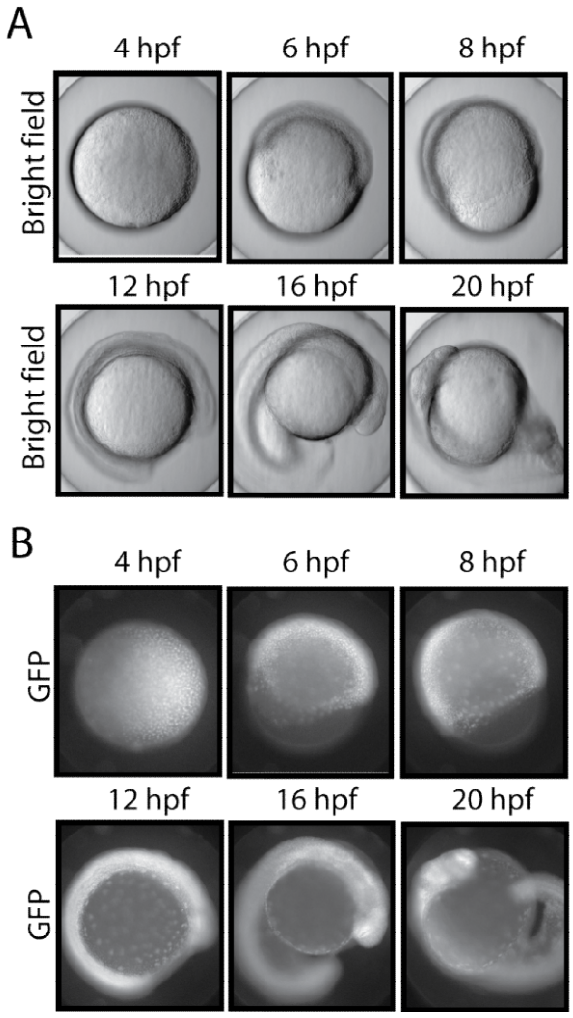



**Fig 3.2.12 H2A:gfp zebrafish embryos we used to optimize GFP imaging of live zebrafish embryos.**

The embryos did survive overnight incubation with the fourteen images taken in both UV (using a GFP filter) and bright field every 30 minutes. However UV illumination of the zebrafish embryos lead to irritation of the embryos. In turn leading to a dramatic increase in zebrafish movement from 16hpf. This made clear imaging of the developing eye difficult.

To immobilise the embryos they were dechorionated and embedded directly in the low melting point agarose. However, this almost invariably leads to the deaths of the embedded embryos **(Fig. 3.2.13)**.

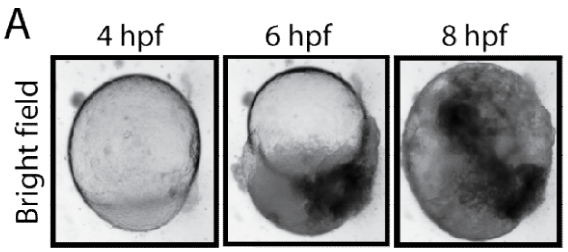



**Fig 3.2.13 Incubation of embryos in 2% low melting point agarose was invariably lethal if they were dechorionated at the sphere stage of development.**

To address the problem of movement the dechorionated embryos were embedded later in development (upon the completion of gastrulation ; 10.5 hpf) using methyl cellulose at a concentration of 0.8% and anesthetised with Tricaine. Without their chorion zebrafish embryos are extremely vulnerable to increases in pressure. To combat this embryos were loaded into the viewing chamber on a minimal amount of methyl cellose. This allowed the presence of air spaces around the media within the viewing chamber, reducing the pressure. The methyl cellose drop was held in place by surface tension **(Fig. 6.3.1)** . These steps minimized the disruption in development and allowed the embryos to develop overnight **(Fig. 3.2.14 and 3.2.15)**.

## 3.2.8 The effect of *zsmoc1* depletion on the developing zebrafish eye

To investigate the effect of *zsmoc1* knockdown on the establishment of the eye field, a rx3:gfp reporter line (Rembold et al. 2006) was used in conjunction with injection of zsmoc1tMO. The dosage of zsmoc1tMO was titrated to minimise the severe dorsalised C2-C4 phenotype reported in **Fig. 3.2.1 A** and in an attempt to recapitulate the phenotype reported in Abouzeid et al. (2011), which was microphthalmia and an associated coloboma. *rx3* is the earliest known marker of the eye field and use of the rx3:gfp reporter enables direct observations of the zebrafish eye field using fluorescent microscopy. *rx3* is expressed throughout the developing eye up to 30 hpf by which time most of the major morphogenic events of eye development have been completed **(Fig. 1.2.1)**. Real time image capture under UV light allows the accurate measurement of the size of the eye field throughout development.

The results shown are from biological replicates, specifically two independent sets of injections but under identical conditions. The analysis showed that the eye field is expanded in *zsmoc1* morphants, with both of the measured axes (a & b; **Fig. 3.2.14 A *i.* and *ii.***) being significantly extended**.** Overall there was no evidence of microphthalmia or coloboma at later stages **(Fig. 3.2.14 A v*i.*)**. The optic cup was larger at 18 hpf in one experiment **(Fig. 3.2.14 A *i.*)** but this was not replicated in the bright field measurements in **Fig. 3.2.15 A *i.*** . All other measurements of eye size are not significantly different **(Fig. 3.2.14 A *i. ii.* and *iii.*)**.

To control for any global effect on embryo size caused by morpholino injection or inconsistent orientation of embryos, the entire cross-sectional area of the embryo **(area d; Fig. 3.2.14 B *ii.* and *iii.*)** was measured along with the width of the embryo head **(axis e; Fig. 3.2.14 B *i.* and *iii.*)**. There was no significant difference in the length of axis e between control and zsmoc1tMO injected embryos, indicating that the embryos were well orientated. There was a slight reduction in the total cross-sectional area of the embryos treated with zsmoc1tMO at both 18 and 22 hpf, which may be indicative of a minor misalignment in the anterior posterior axis or may be related to zsmoc1tMO and represent a milder form of the anterior posterior shortening observed in C2 - C4 embryos. Such a reduction cannot account for the results obtained for the optic cup at 18 and 22 hpf as they result in a smaller proportional optic cup size for both a smaller embryo or poorly aligned embryo, confirming that zsmoc1tMO does not induce microphthalmia.

zsmoc1tMO injections also appeared to have little effect on *rx3* expression. There may be a slight up-regulation of *rx3* at 14 hpf in *zsmoc1* morphants although the differences are only statistically significant for the mean pixel intensity **(Fig. 3.2.14 A *iv.*)**. At all other time points there is no statistically significant difference in pixel intensity of zsmoc1tMO and control MO injected zebrafish **(Fig. 3.2.14 A *iv.* , *v.* and *vi.*)**.

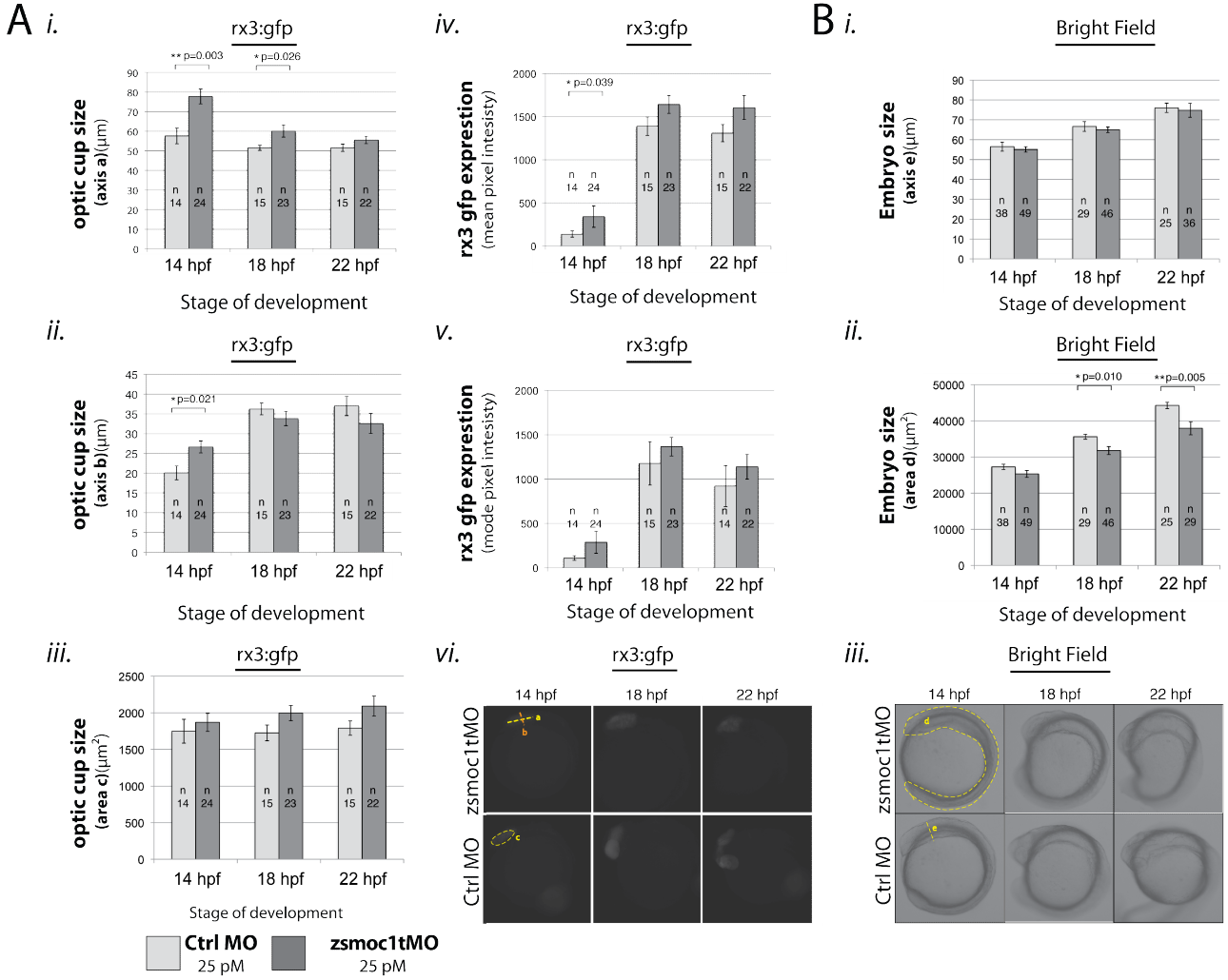



**Fig 3.2.14 The size of the optic cup is unaffected in *zsmoc1* morphants in rx3:gfp reporter lines.**

***(A)(i)(ii)*** The eye field is enlarged at 14 and 18 hpf in both axis a and b, However area c remains unaffected ***(A)(iii)***. ***(A)(iv)*** *zsmoc1* depletion leads to a slight up regulation of rx3:gfp expression at 14 hpf. **(B)** Zsmoc1tMO treated show a reduction in embryo size estimated by area d **(B)(ii)** of the embryo at both 18 and 22 hpf. This reduction was not replicated in embryo size estimate by measuring axis **e (B)(i).**

Example axis are illustrated in **A(vi; gfp) B(iii; brightfield)** for more details consult 6.3.1.4.
n=number of embryos
*All data presented as mean (± s.e.m) and the p-values were calculated using a type 3 non parametric ttest.

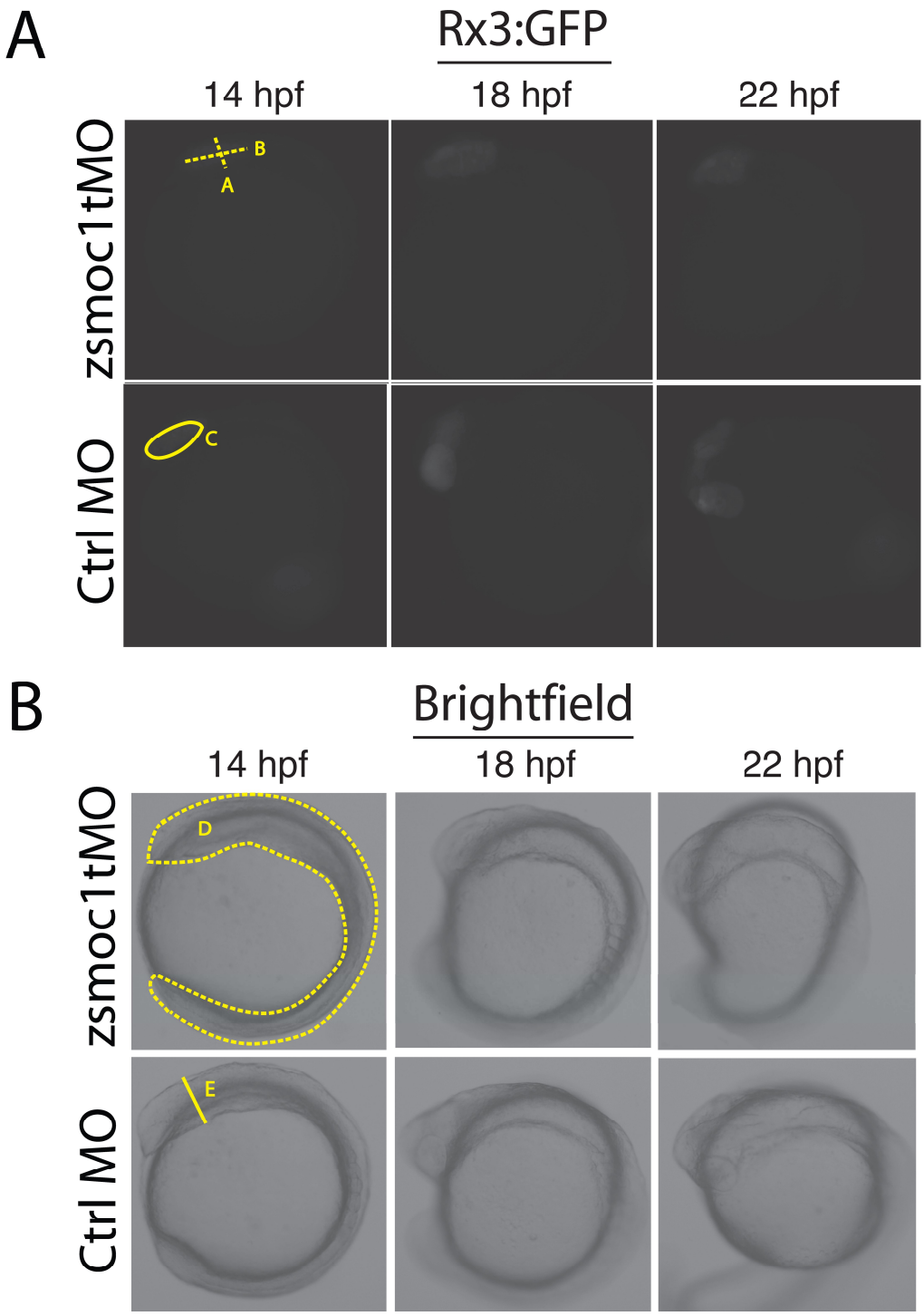



**Fig 3.2.14b The size of the optic cup is unaffected in *zsmoc1* morphants in rx3:gfp reporter lines.**

Example axis are illustrated in **A(vi; gfp) B(iii; brightfield)** for more details consult 6.3.1.4. n=number of embryos

# BMP signalling in the developing zebrafish eye

To visualise the bmp activity gradient formed within the developing eye, we used live imaging of BRE (Bmp response element):gfp transgenic embryos (Collery & Link 2011) injected with control or *zsmoc1* morpholinos. The embryos were assessed at 22 hpf, which was one hour after the BRE:gfp gradient became visible (data not shown). *zsmoc1* morphants show a small but significant decrease in the size of axis g **(Bright field; Fig. 3.2.15 A)**. However, this difference was not observed in the rx3:gfp embryos indicating that the result is spurious and that the eye field is not reduced by *zsmoc1* depletion **(Fig. 3.2.15 A *i., ii.* and B *iii* )**.

As with the rx3:gfp embryos, zsmoc1tmo treatment results in smaller embryos. However, the differences are only statistically significant at 18 hpf **(Fig. .2.15 B *ii.*)** and since they are also present in the measurements of axis e **(Fig. 3.2.15 B *i.*)** this may be a product of embryo misalignment.

While there was no difference in eye size there was a significant difference in the gradient of BRE activity. *zsmoc1* depletion made no significant difference in the absolute size of the BRE activity gradient, either in terms of the width (axis a and b ; **Fig. 3.2.15 A *iii. and vi.)*** or the cross-sectional area ( area c ; **Fig. 3.2.15 A *iv. and vi.)***. However *zsmoc1* morphants had a higher ratio of BRE fluorescence to eye size, suggesting that in *zsmoc1* morphants the dorsoventral gradient extends through a greater volume of the eye field.

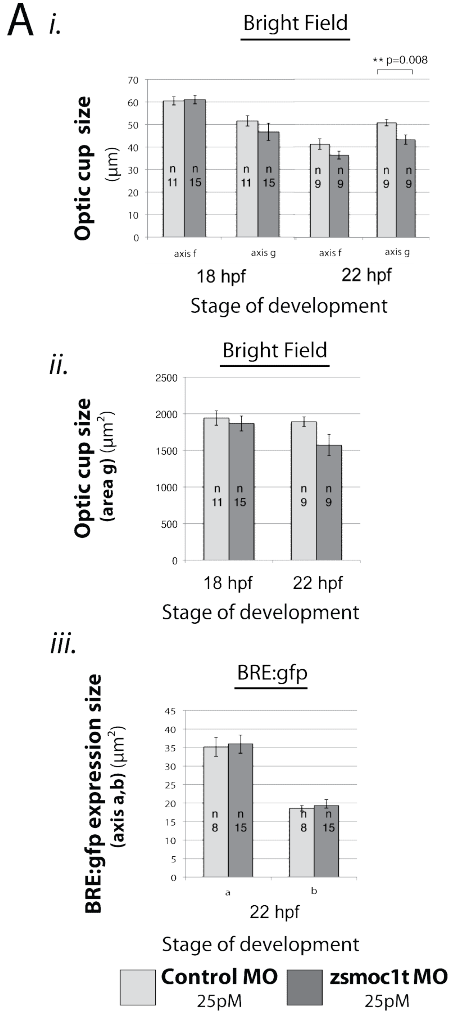



**Fig 3.2.15 Upregulation of bmp signalling in the dorsal RPE of *zsmoc1* morphants.**

**(A)**The effect of *zsmoc1* depletion on the optic cup development. **A(i.)(ii)** With the exception of axis g at 22 hpf the overall size of the eye field is not significantly different between *zsmoc1* and control morphants**. A(iii)(iv)** The size of the bmp gradient remains unaffected by *zsmoc1* depletion. However the total ratio of the eyefield taken up by BRE expression is greater in *zsmoc1* morphants. **(A)(v)** The expression of the bmp reporter BRE:gfp is increased in *zsmoc1* morphants.
**(B)** At 18 hpf *zsmoc1* morphants display a significantly reduced size, as measured by area d and axis e.

Example axis of interest and regions of interest are illustrated in **A(vi; gfp) B(iii; brightfield)** for more details consult 6.3.1.4.
n=number of embryos.
* All data presented as mean (± s.e.m) and the p-values were calculated using a type 3 non parametric ttest.

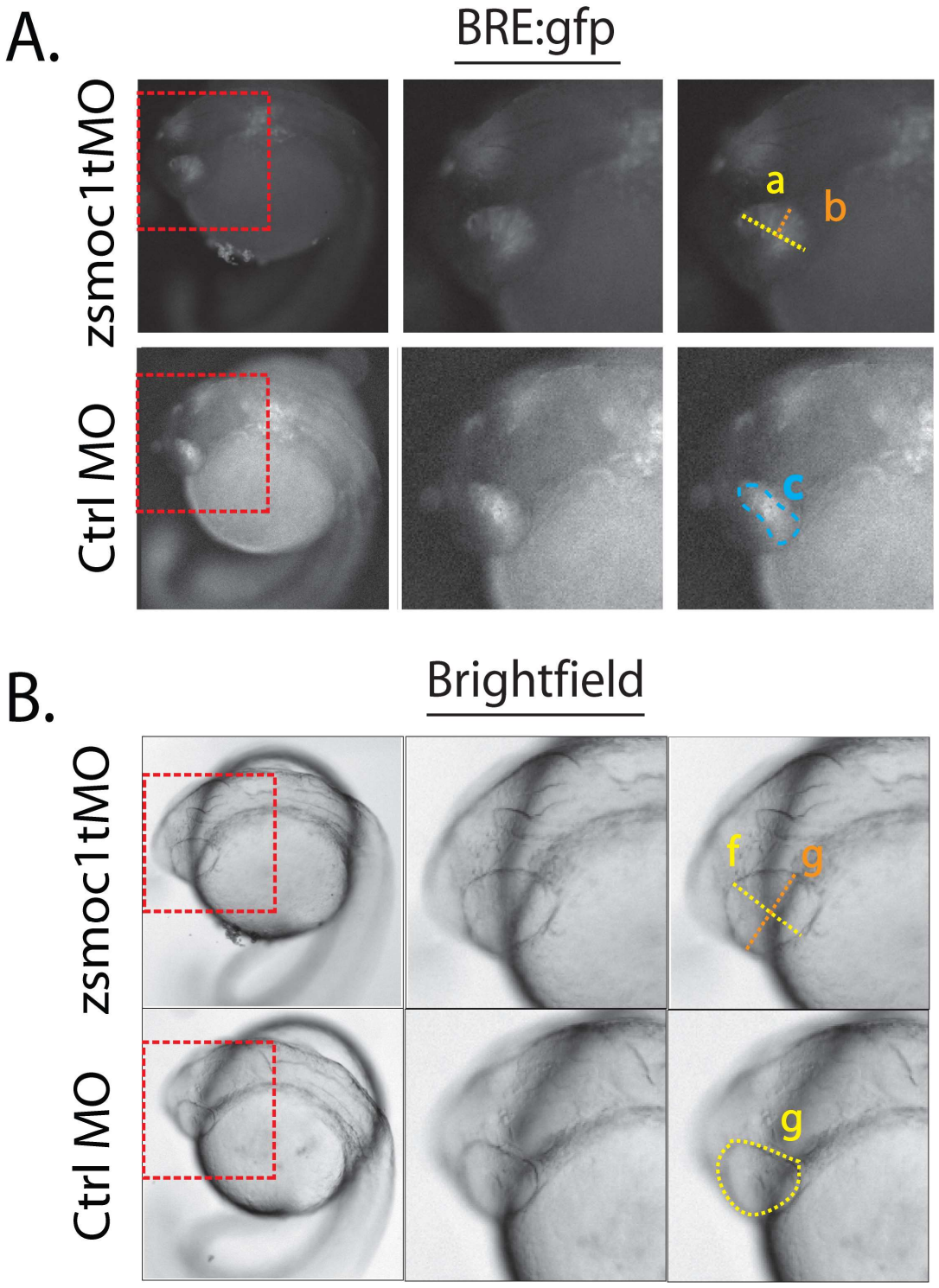



**Fig 3.2.15b Upregulation of bmp signalling in the dorsal RPE of *zsmoc1* morphants.**

Example axis of interest and regions of interest are illustrated in **A(vi; gfp) B(iii; brightfield)** for more details consult 6.3.1.4.

Most notably, *zsmoc1* morphants displayed a significant up regulation in BRE activity. The mean pixel intensity in *zsmoc1* morphants is 2.5 times that of embryos injected with control MO **(Fig. 3.2.15 A *v. and vi.*;** p= 1.6 x $10^{-6}$ ). The mode intensity is twice that of wild type embryos **(Fig. 3.2.15 A *v. and vi.*;** p= 2.9 x $10^{-5}$ ) demonstrating that *zsmoc1* acts as a bmp antagonist influencing the dorsal ventral gradient in the optic cup.

## 3.2.10 *zsmoc1* morphants exhibit and a laterally expanded dorsoventral p-smad gradient at 60% epiboly(6hpf)

To determine whether *zsmoc1* directly affects the shape of the dorsal ventral bmp gradient of gastrulating zebrafish embryos whole mount immuno-histochemistry was performed against p-smad(1/5/8). The embryos were fixed at 60% epiboly because this is the time point where the epibolic defects of *zsmoc1* morphants become apparent (David Sexton 2010). The p-smad gradient was visualised with horseradish peroxidase (HrP) staining before being quantified using Optical Projection Tomography (OPT).

The 3d projection of the zebrafish gastrulating embryos was optically sectioned in 2 planes. To produce the quantitative measure of gradient size and identity the region of interest was defined as the part of the gradient which had an signal intensity of greater than 200. The gradient was then analysed with two parameters. The extent of the gradient was measured by quantifying the size of the

region of interest and the mean pixel intensity was measured within the region of interest.

The lateral section allowed the measurement of the Anterior pole (AP) gradient. Which is the gradient that controls the dorsoventral patterning of the developing zebrafish (Solnica-Krezel & Sepich 2012). The analyses of the lateral sections showed there was no statistically significant difference in the size of the AP gradient in *zsmoc1* morphants. **(Fig. 3.2.16 A and B).**

The embryos were also section along the AP axis as this allowed the visualisation of the lateral p-smad gradient. It is this part of the p-smad gradient that controls the dorsal migration of lateral mesoderm during convergence of the gastrulating zebrafish embryo (von der Hardt et al. 2007). The AP sections showed that the size of the lateral gradient was significantly expanded in *zsmoc1* morphants **(Fig. 3.2.16 C and D).** Which is consistent with an antagonistic function of zsmoc-1.

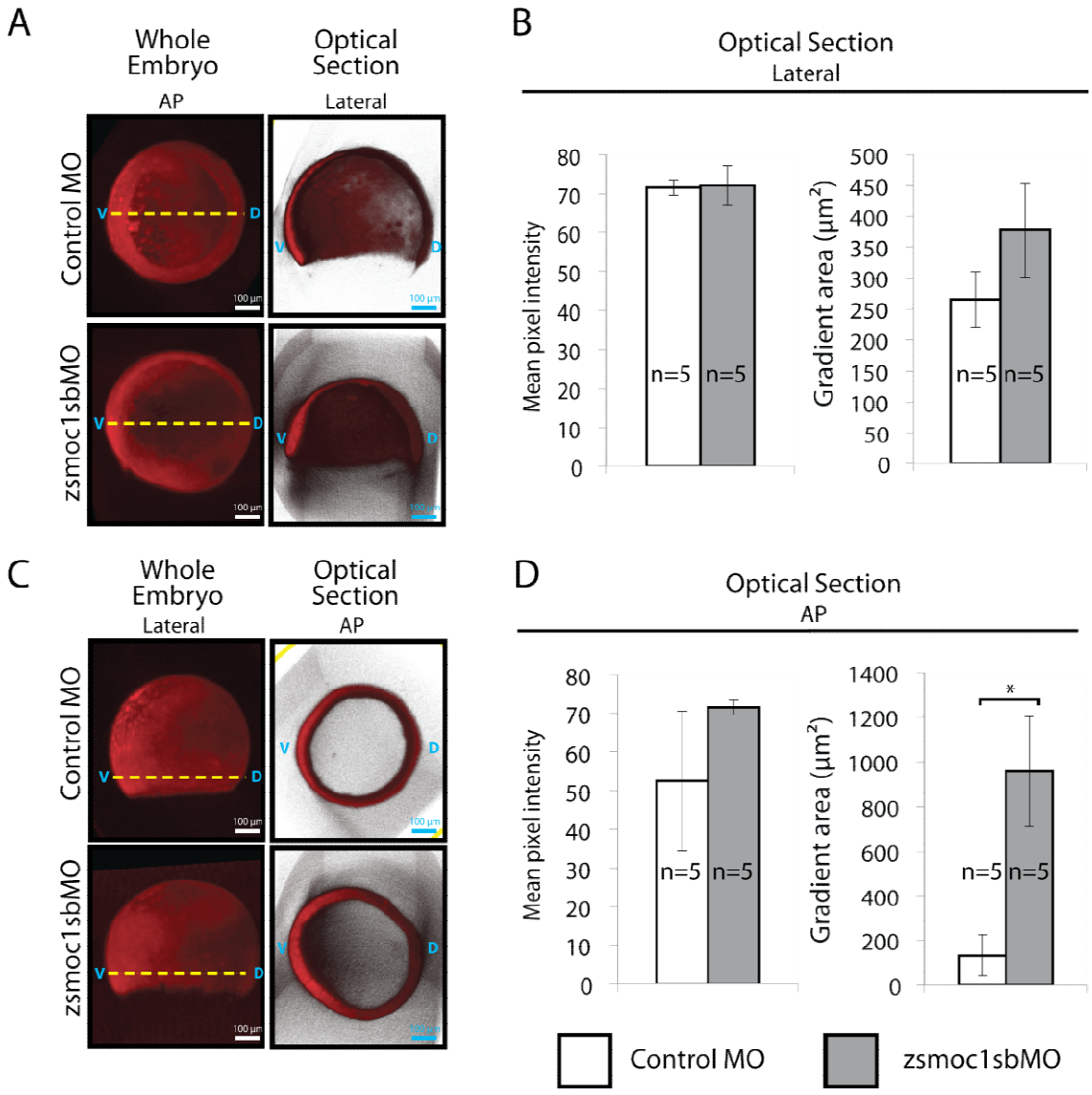



**Fig 3.2.16 *zsmoc1* constrains the sagittal p-smad gradient of zebrafish at 60% epiboly.**

OPT imaging was used to produce three-dimensional scans of whole mount immuno. The p-smad gradient was visualised using anti-p-smad 1/5/8 and HR staining.
**(A)** Images illustrating the optical sectioning of the whole mount immunos. The first, column shows a 3-D reconstruction of the HRP staining of the embryo from a vegetal view. The second column depicts an example of a lateral optical section representative embryo. **(B)** The gradient lateral p-smad gradient is unaffected by zsmoc1 depletion.
**(C)** Images illustrating the VP-AP sectioning of the zebrafish embryos. The first column depicts 3-D reconstructions of embryos from the lateral plane. The second column defects optical sections taken in the sagittal plane. **(D)** While the intensity of the gradient is not significantly different in smoc1 morphants the gradient is significantly larger in zsmoc1 morphants.
Images were initially constructed and optically sectioned using Biotronics V4.0 then exported as Tiff files and analysed in Fiji

V= Ventral D= Dorsal AP= Animal Pole VP= Vegital Pole

The axes of the optical section is denoted by the yellow dashes.

Results displayed as the mean± s.e.m. Statistical significance was calculated using a nonparametric two-tailed students t-test. *=p<0.05 **=p<0.001.

# 3.3 Discussion

## 3.3.1 zsmoc-1, an antagonist of BMP signalling?

When I started this work the prevailing model was that smoc-1 acted as a simple antagonist of bmp signalling. This model is largely based on one paper, Thomas et al. (2009) which demonstrated that in *Xenopus* embryos *xsmoc1* morphants display the ventralised phenotype characteristic of a bmp antagonist and that *xsmoc1* overexpression resulted in reduced smad phosphorylation. They also showed that the ventralisation was bmp-dependent by showing that it was possible to rescue *xsmoc1* morphants with the expression of constitutively active type one bmp receptor.

Yet the results presented in this chapter have demonstrated that this is not a complete picture, at least not in zebrafish. While multiple lines of evidence all implicate *zsmoc1* in bmp signalling the results are largely contradictory with some pointing to bmp acting as a classical antagonist **(Fig. 3.1.4 F; Fig. 3.2.15; Fig. 3.2.16)** while others point to an agonistic function **(Fig. 3.2.6)**.

The evidence that zsmoc-1 acts as an antagonist of bmp signalling is threefold. 1). *Zsmoc1* morphants have universally up-regulated smad phosphorylation at the sphere stage of embryonic development (David Sexton 2010) 2). *zsmoc1* morphants display a laterally expanded p-smad gradient at 60%

epiboly **(Fig. 3.2.16 C and D)** and 3). Lower doses of zsmoc1tMO result in an up regulation of bmp signalling within the optic cup, as shown by the increase in the intensity of the BRE:gfp signal **(Fig. 3.2.15 A *v.* and *vi.*)**.

However overexpression of *hSMOC1* led to a ventralised phenotype. The neural markers were down regulated while the ventral ectoderm markers were upregulated**(Fig. 3.2.5 A and *B*).** Which is the pattern you would expect from a bmp agonist and indeed this was the pattern of gene expression observed with the injection of *bmp2b* mRNA **(Fig. 3.2.3 A; B and** C).

One of the most unexpected results was the use of RT PCR to demonstrate that the *zsmoc1* morphant phenotype observed in the preliminary work before this project was not standard dorsalisation. Instead of showing up regulation of the neural marker genes and concurrent down-regulation of ventral ectoderm marker genes one would expect in dorsalised embryos, the *zsmoc1* morphants showed universally down regulated gene expression **(Fig. 3.2.4 A and B)** which conclusively show that the *zsmoc1* morphants are not dorsalised. This still raises the question as to why the morphants display the C2-C4 phenotype **(Fig. 3.2.1)**?

One explanation is that the embryonic defects are the result of a general failure of differentiation of the zebrafish embryo. Thus resulting in poorly differentiated tissues that fails to recapitulate the expression patterns of ventral and neural ectoderm. This model is supported by the phenotype of OAS patients. Both of the key defects in OAS (anophthalmia and syndactyly) are theorised to be due to failure of differentiation (Verma & Fitzpatrick 2007; Tonkin 2009). A large proportion of the causative genes (for example *OTX2* and *RAX*) for anophthalmia/

microphthalmia are expressed exclusively in the RPE during eye development, leading Verma & Fitzpatrick to propose that anophthalmia/ microphthalmia is caused by the failure of retinal to differentiation. Resulting in the failure of learners induction and the eventual regression of the eye field (Verma & Fitzpatrick 2007). Syndactyly is known to be due to failure of differentiation. More specifically incomplete differentiation of the digital rays (Peter D. Burge 2009; Tonkin 2009).

An alternative explanation is that zsmoc-1 is involved in controlling the shape of the bmp gradient that regulates dorsal migration of the lateral mesoderm in convergence. This has already been observed in *knypek* mutants. *Knypek* is a zebrafish homolog of *Glypican 4*(*GPC4*). *Knypek* null embryos display C2 phenotypes but the embryos aren't dorsalised. Instead the defects are attributable to the inability of *knypek* null cells to migrate along the lateral component of the dorsal ventral p-smad gradient (Topczewski et al. 2001; Kondo 2007). This is supported by the whole mount immuno OPTs which show no defect in the animal pole bmp gradient **(Fig. 3.2.16)**. This is the part of the dorsoventral gradient which determines the identity of the mesoderm, i.e. whether it differentiates neural or ventral mesoderm in the zebrafish embryo, and is the part which would alter the expression of the dorsal ventral markers. This is in contrast to the lateral gradient, which is both expanded and extended dorsally. The lateral component of the dorsoventral gradient controls the migration of the lateral mesoderm, but not the eventual cell fate of the mesoderm (Topczewski et al. 2001; von der Hardt et al. 2007). This accounts for the observation that defects in *zsmoc1* morphants only become observable at 50% epiboly (David Sexton 2010), which is when convergence begins. An interaction with knypek would also be in line with the

known function of *Pentagone* (the *Drosophila* homologue of *smoc1)*, which has been shown to be dependent on a direct interaction with *Dally* (the *Drosophila* orthologue of the *knypek*; Vuilleumier et al. 2010).

The final possibility is that the results of Thomas et al. (2009) were fundamentally flawed. Given the current controversies surrounding the use of morpholinos in zebrafish, namely the surprisingly poor correlation between morphant phenotypes and those of genetic knockouts(Kok F. et al. 2014), it is worth reassessing the results reported by Thomas et al. Unfortunately there has been relatively little work on assessing the extent of the problem in *xenopus*, with most of the prevailing assumptions derived from the results obtained in zebrafish (Eisen & Smith 2008: Wang, et al. 2015.). Given the standards of the time the work was conducted with the necessary controls. The paper did target multiple morpholinos to the same gene, however given the results in zebrafish this cannot be considered definitive. Fortunately Thomas et al. conducted multiple other experiments that did not rely on morpholino data. Most notably the injection of *xsmoc1* mRNA into *xenopus* embryos resulted in the upregulation of ventral markers and *xsmoc1* mRNA can block the dorsalisation produced by the direct injection of bmp2 protien (Thomas et al. 2009). Both these experiments provide strong support for an antagonistic function of smoc-1 and cannot be accounted for by morpholino toxicity.

## 3.3.2 A model for the feedback between zsmoc-1 and bmp signalling

It is clear that zsmoc-1 neither acts exclusively as an bmp antagonist or agonist. It is entirely possible that the action of zsmoc-1 is context dependent, as is the case for Schnurri (Dai et al. 2000; Arora et al. 1995). However, a more parsimonious explanation is that zsmoc-1 acts to control the shape of the gradient. This is why we have proposed that the function of *zsmoc1* is conserved to *Drosophila* and that zsmoc-1 acts on the bmp gradient in much the same way that Pentagone has been shown to act on the Dpp gradient (Vuilleumier et al. 2010; Hamaratoglu et al. 2011).

Like *Pentagone*, *zsmoc1* is repressed by its own downstream target. *bmp2b* overexpression results in the repression of *zsmoc1* expression, and this repression is dosage dependent **(Fig. 3.2.3)**. While *bmp2b* morphants have higher levels of *zsmoc1* expression than the embryos injected with control MO **(Fig. 3.2.2)**. Reciprocal repression is the characteristic expression pattern of a feedback loop which is important in signalling cascades as it allows the generation of robust gradients, allowing morphogen gradients to compensate for genetic disruptions (if gradients are not dynamically regulated the of majority of signalling ligands would have monoallelic disorders. as any single null mutation would result in a significantly reduced pool of available ligand) and environmental changes (e.g. temperature; Plouhinec & De Robertis 2009). They also allow gradients to be

modified dynamically during development and impart properties such as scalability to the gradient (Hamaratoglu et al. 2011).

Under our proposed model, *zsmoc1* acts in a feedback loop with bmp and smad1/5/8. In this model bmp ligand leads to smad phosphorylation while simultaneously suppressing the expression of *zsmoc1*. While we have assumed that this suppression is mediated via p-smad, there is no direct evidence for this and it is equally possible that *zsmoc1s* action is dependent on non canonical bmp signalling (e.g. MAPK pathway; Verheyen 2007; Sieber et al. 2009). zsmoc-1 promotes the migration of bmp ligand, but the precise mechanism of this remains to be elucidated. In *Drosophilae,* however, the regulation of Dpp signaling by Pentagone is endocytosis dependent which was demonstrated by the rescue of the bmp gradient with the knockdown of *Thickveins* (Vuilleumier et al. 2010).

Under RMT, the formation of the gradient is dependent on transcytosis. zsmoc-1 would promote transcytosis which would spread the pool of available ligand resulting in a broader less concentrated bmp gradient (Schwank et al. 2011). Under the RED model, zsmoc-1 would inhibit the endocytosis of bmp. This endocytosis is critical as smad phosphorylation is dependent on the correct intracellular localisation (Schwank et al. 2011). The internalisation of bmp/bmpr complex is also critical to its degradation, so if zsmoc-1 acts by inhibiting endocytosis the available pool of ligand would be increased and extend the bmp gradient(Schwank et al. 2011).

In practice this would lead to smoc-1 being expressed only at a distance from the source of bmp ligand. This indeed is the expression pattern observed in dorsoventral retinal eye patterning, with bmp being expressed in the dorsal retina and *zsmoc1* expression restricted to the ventral retina by the optic fissure (French et al. 2009; Abouzeid et al. 2011). For the dorsoventral patterning of the zebrafish embryo, the expression domain is less certain. While q-PCR has demonstrated that *zsmoc1* is expressed during gastrulation, the expression levels are relatively low and the expression domains have not been established **(Fig. 2.2.6)**. Given that *zsmoc1* morphants only display perturbed gradients in the sagittal optical sections **(Fig. 3.2.16),** it is likely that *zsmoc1* is only expressed laterally and is not involved in regulating the sagittal gradient.

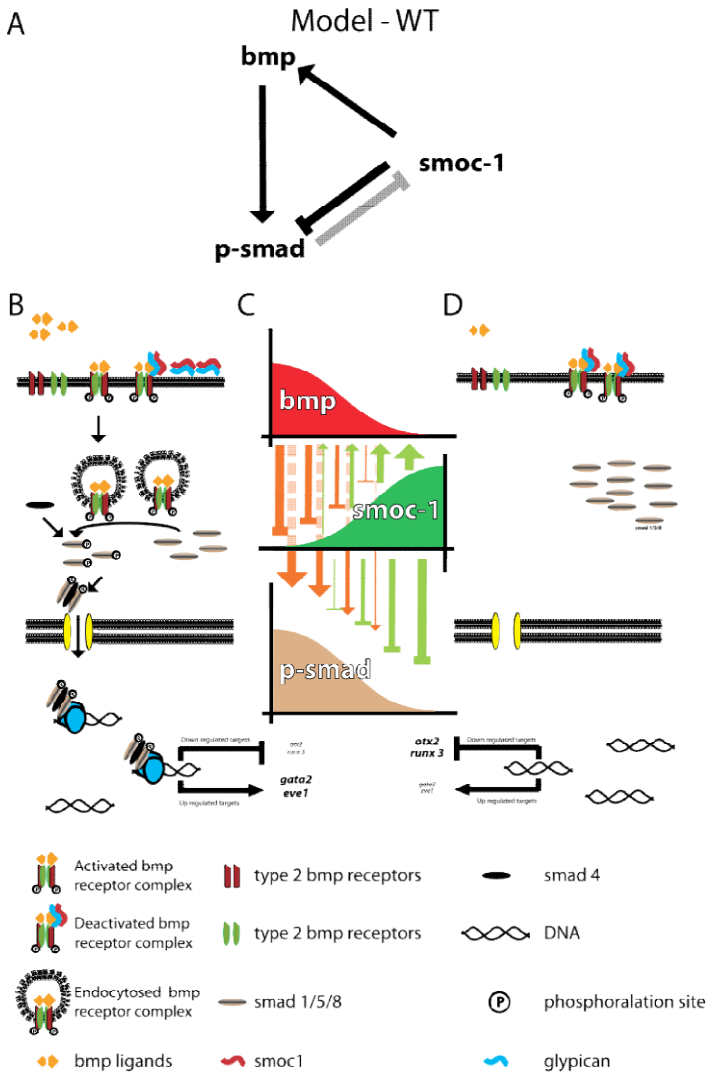



**Fig. 3.3.1 Suggested model of zsmoc-1 action**

**(A)** smoc-1 acts in a feedback loop with bmp signalling. bmp initiates smad phosphorylation. smoc-1 both promotes the formation of bmp gradients by inhibiting ligand degradation/ migration and controls smad phosphorylation by maintaining the pool of available BMP. **(B; D)** In turn BMP down regulates smoc-1 action by inhibiting *zsmoc1* expression.

**(B)** At the bmp source the available pool of bmp is high and therefore smad signalling is high. The repression of smoc1 expression by bmp will result in the establishment of an opposing smoc-1/glypican gradient. **(D)** Away from source any bmp will be inhibited from being endocytosed. This will reduce smad phosphorylation while promoting gradient migration. Effectively buffering the bmp gradient from perturbations.

The transparent inhibition symbol of smad inhibition of smoc-1 represents the speculative nature of this interaction. Though there is evidence that bmp levels control smoc1 expression it could be through non-canonical downstream signalling.

---

The injection of *bmp2b* mRNA results in universal expression of *bmp2b*. Under my model of development, this will result in the repression of *zsmoc1* and the up-regulation of p-smad signaling. Normally a down regulation of *zsmoc1* would result in the contraction of the bmp gradient, however injected mRNA is ubiquitously expressed (Finckbeiner et al. 2011; Skromne & Prince 2008). Under the RED model, endocytosis would be up-regulated by the absence of smoc-1, up-regulating p-smad signalling even further. The p-smad goes on to upregulate

ventral ectoderm markers (e.g. *eve1* and *gata2*), while reducing the expression of the neural markers (e.g. *otx2* and *runx3*; **Fig. 3.3.2)**.

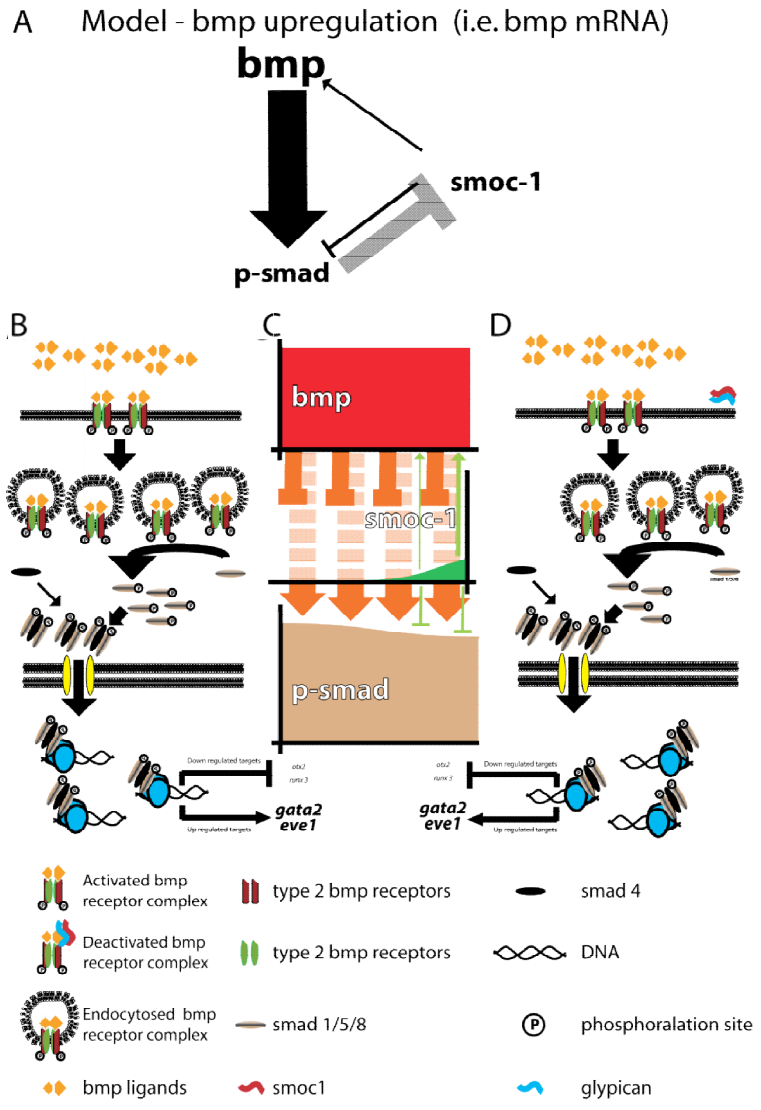



**Fig. 3.3.2 Suggested model of zsmoc-1 action - in embryos where bmp signalling has been up regulated with *bmp2b* overexpression.**

Injecting embryos with *bmp2b* mRNA results in a up-regulation of bmp signalling **(A)**. Which in turn increases the levels of p-smad. psmad signalling then goes on to inhibit *zsmoc1* expression. This would both decrease bmp migration and increase smad phosphorylation. However as *bmp2b* mRNA is ubiquitously expressed there is no need for bmp ligand to migrate. Which means the net result is a massive up-regulation in BMP signalling **(C).**

This results in the universal up-regulation of genes under the positive control of bmp and the universal downregulation of those genes under the negative control of bmp. Which in the case of the ventral bmp gradient in the test gastrulating embryo would lead to the up-regulation of *gata2* and *eve1* and down-regulation of *otx2* and *runx3***(C).**

bmp2b is the master regulator of the dorsoventral gradient during gastrulation in zebrafish, and initiates the expression of both bmp4 and bmp7 (Kishimoto et al. 1997; Kondo 2007). When *bmp2b* is down regulated with the injection of bmp2b MO, all smad phosphorylation is inhibited (von der Hardt et al. 2007). The removal of bmp signalling results in an up-regulation of *smoc1*. The

smoc-1 would encourage the migration of bmp ligand inhibiting smad phosphorylation, though in this case there is little/no bmp ligand left for smoc-1 to act on **(Fig. 3.3.3)**.

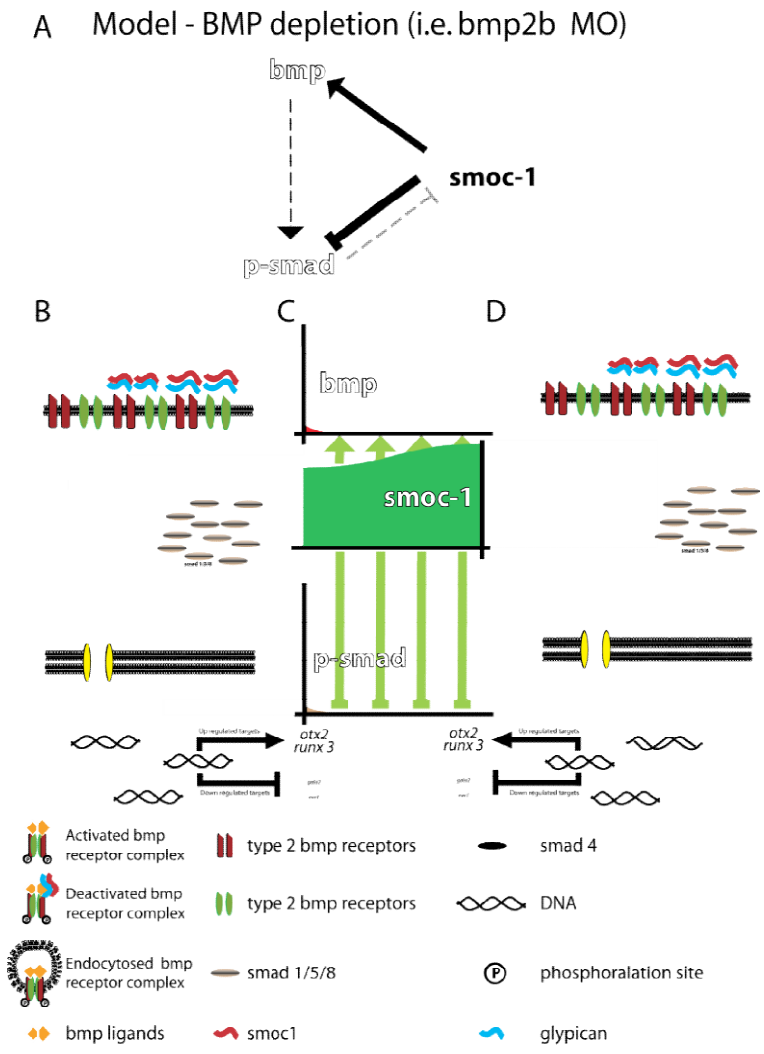



**Fig. 3.3.3 Suggested model of zsmoc-1 action - in embryos where bmp signalling has been ablated with *bmp2bMO***

The injection of bmp2b leads to the almost absence complete absence of bmp ligands **(A;C)**. With no bmp there is no smad phosphorylation. This in turn releases the repression of *zsmoc1* resulting in increased expression smoc1**(C)**. However as there is almost no free bmp = to act on, the increase in zsmoc-1 levels has little effect on BMP signalling.

The advantage of this model is that it explains why *hSMOC1* overexpression acts as a bmp agonist, while when *zsmoc1* is depleted it shows the phenotype expected from a bmp antagonist. Under this model, the ubiquitous *smoc1* expression caused by mRNA injection would lead to a broader bmp gradient. It

would lead to lower p-smad phosphorylation levels near the source of the gradient but there would still be significant phosphorylation. The more dramatic effect would be on the distal end of the dorsal ventral gradient. bmp ligand would make it further into the dorsal axis of the embryo, ventralising it and resulting in an overall increase in the ventral ectoderm markers at the expense of the dorsal neural markers **(Fig. 3.2.7 ; Fig. 3.3.4).**

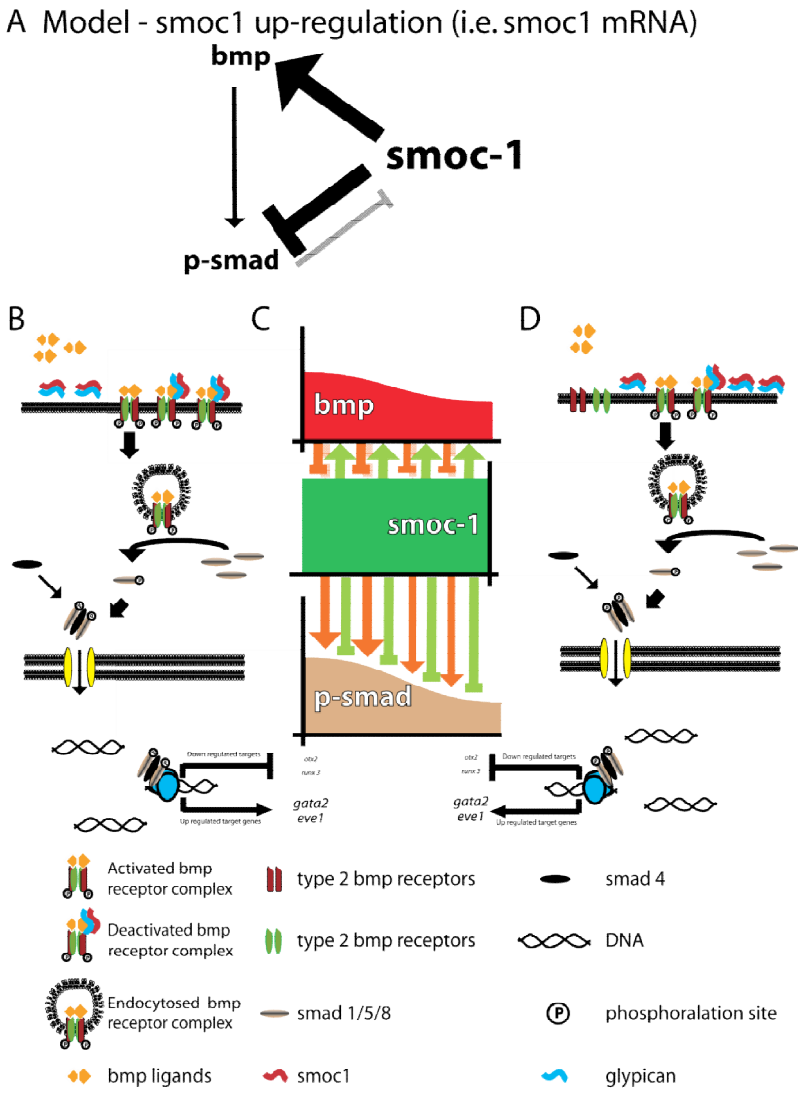



**Fig. 3.3.4 Suggested model of zsmoc-1 action - in embryos injected with *hSMOC1* mRNA.**

Overexpression of SMOC1 results in a significantly greater and BMP migration, but inhibits smad phosphorylation**(A)**. This results in a reduced pool of bmp near the gradient source as it has migrated further from source. This results in a reduction in the level of p-smad signalling resulting in a lower relative levels of positively regulated bmp targets and relatively higher levels of negatively regulated BMP targets. Though the absolute levels of positively regulated genes remain higher. The distal end of the bmp gradient has significantly more bmp signalling. Resulting in an increase in the level of p-smad signalling. Leading to an up-regulation of positively regulated genes and down-regulation of negatively regulated genes.

Overall the level of expression of the positively regulated genes is greater. During the dorsoventral patterning of the embryo *gata2* and *eve1* expression will be up regulated while the expression of *otx2* and *runx3* are down-regulated.

How does this model account for zsmoc-1 depletion acting like a bmp antagonist and up regulating bmp signalling **(Fig. 3.1.4; Fig 3.2.15)**? The RED model predicts that smoc-1 ablation will remove the inhibition on endocytosis, while the RMT model predicts that the gradient migration will be inhibited. Both models will result in a buildup of bmp ligand near to the source (Schwank et al. 2011; Kicheva & González-Gaitán 2008) resulting in a localised up-regulation of p-smad signalling but a complete ablation of distal p-smad signalling. This theory does account for the universal down-regulation of the dorsal and ventral marker genes, as it makes no clear prediction on the levels of gene expression. Under my model the ventral side of the embryo be ventralised and the dorsal side of the embryo being dorsalised **(Fig. 3.3.5)**.

It is more likely that zsmoc-1 is not involved in the dorsal ventral patterning in gastrulation; whole mount immunos showed that *zsmoc1* morphants only displayed defects in the lateral p-smad gradient **(Fig. 3.2.16)**. This gradient is not involved in the establishment of the dorsal ventral cell fate but is involved in regulating cell migration through convergence and anterior posterior extension (Solnica-Krezel & Sepich 2012). *hSMOC1* overexpression does affect the dorsoventral axis because it is a ectopically expressed, allowing it to act on the AP component of the gradient. This part of the p-smad gradient is involved in establishing cell fate (Solnica-Krezel & Sepich 2012).

What this model does not address is whether smoc-1 acts on other signalling pathways. There is evidence to support this; *xsmoc1* acts before p-smad, as *xsmoc1* mRNA injections could not rescue the phenotype of embryos with constitutively active p-smad (Thomas et al. 2009). Some of this makes little sense, as smoc-1 is an extracellular protein localised to the basement membrane (Gersdorff et al. 2006), suggesting that smoc1 should act on or upstream of the bmp receptors. Thus, it is possible that smoc1 acts through the non-canonical bmp pathway (Hu et al. 2004; Hartung et al. 2006), which is smad independent and involves signalling via the MAPK pathway. Consistent with this model, excess Smoc-1 has been shown to reduce Smad signalling and increased p-Erk levels in rat tissue culture (Dreieicher et al. 2009).

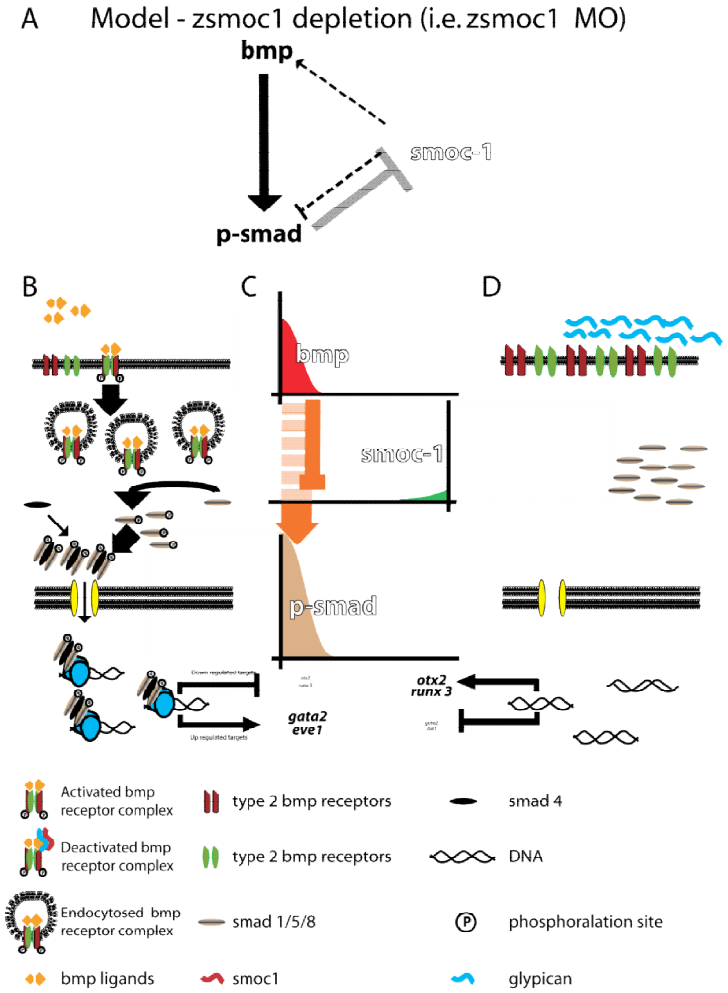



**Fig 3.3.5 Suggested model of zsmoc-1 action - in embryos injected with zsmoc1MO underRED**

Zsmoc1 depletion leads to increased endocytosis. Which in turn leads to an up-regulation of smad phosphorylation **(B)**. However this simultaneously depletes the available pool of bmp ligand. Resulting in a contracted bmp gradient **(C)**. The end result is a more intense but greatly contracted psmad gradient. During gastrulation ventral genes will be up regulated when compared to wild-type. However, as they are further away from the bmp source dorsal genes will be up-regulated **(D)**.

## 3.3.3 *zsmoc1* depletion does not result in a morphological change in eye development

Live imaging of the rx3:gfp line failed to show any reduction in the size of the eye field of embryos injected with zsmoc1tMO **(Fig. 3.2.16 A *i. ii.* and *iii.*)**. In fact the eye field was expanded at 18 hpf **(Fig. 3.2.16 A *i.* and *ii.*).** Whilst this expansion of rx3:gfp expression remains unexplained, it is consistent with the down-regulation of bmp signalling. The eye field in teleost fish is derived from the telencephalon, and bmp2b inhibits the expression of *rx3* and prevents its differentiation into the eye field (Bielen & Houart 2012). A down-regulation of bmp2b signalling would result in an expansion of the eye field and up-regulation of *rx3*. This was observed in the increase in the mean pixel intensity of the eye field rx3:gfp**(Fig. 3.2.16 A *iv.*)**. This would require that zsmoc-1 acts as a agonist of bmp signalling which is inconsistent with other results we obtained. Additionally, bmp2b is not the only signalling molecule involved in the establishment of the eye field. For example wnt signalling is required to regulate the formation of the telencephalon and the establishment of the eye field (Houart et al. 2002; Heisenberg et al. 2001). It is also possible that the expansion of the eye field is attributable to a delay in embryonic development caused by the injection of zsmoc1tMO.

By 18 hpf, *zsmoc1* depletion appears to have no effect on the optic cup. There is no difference in the size or cross-sectional area of the eye fields between

embryos injected with zsmoc1tMO or control MO **(Fig. 3.2.14 and 3.2.15)**. This situation continues to 22 hpf. There were no observed instances of coloboma either. Even after five days of development, both zsmoc1tMO and zsmoc1tMO morphants showed no enrichment of eye defects (microphthalmia/ coloboma and lens extrusion) when compared to control morphants. This is surprising; every other animal model displays eye defects when *Smoc1* is down regulated. Humans with no functional alleles of *SMOC1* develop anophthalmia/microphthalmia (Abouzeid et al. 2011; Rainger et al. 2011; Okada et al. 2011) and mice develop coloboma (Rainger et al. 2011; Okada et al. 2011). *Xenopus* develop anophthalmia when injected with *xsmoc1* targeting MO (Thomas et al. 2009). Though it should be noted that the phenotype only becomes apparent upon unilateral injection of *xsmoc1* MO into one cell of a two cell embryo. Universal down regulation of *xsmoc1* leads to gastrulation defects that arrest embryonic development before the establishment of the eye field (Thomas et al. 2009). Moreover in 2011 Abouzeid et al. reported that *zsmoc1* targeting MO induced microphthalmia and coloboma in zebrafish.

How can the results we obtained be reconciled with those of Abouzeid et al.? The best explanation is MO toxicity or off target effects. Both microphthalmia and coloboma are known to be common phenotypes for morpholino toxicity (Bedell et al. 2011). Morpholinos often trigger the tumour suppressor p53, which leads to apoptosis (Gerety & Wilkinson 2011). In zebrafish, up to 20% of MOs lead to some degree of p53 dependent toxicity (Ekker & Larson 2001) and is notable that while the Abouzeid et al. paper did use RT PCR to demonstrate that their

morpholino did knockdown of *zsmoc1,* they did not take steps to control for off target effects (Abouzeid et al. 2011). They only used one morpholino, they didn't co-inject with p53 MO and they didn't manage to rescue with mRNA.

In contrast, multiple morpholinos were used to knockdown *zsmoc1* expression in this project. In total over the course of preliminary work (David Sexton 2010) and the work presented in this thesis, four separate morpholinos were used which all produced overlapping phenotypes. This is important as it rules out chance off target binding effects. While MOs are 25 bp long, which means the chances of that precise sequence occurring in the genome are very small (The number of combinations of a 25 bp oligo is $4^{25} = 1.126 \times 10^{15}$ so the probability of a precise match at random is $1/1.126 \times 10^{15}$), even a partial match may be enough to produce some binding and knockdown an untargeted mRNA. However the chances of this being the case with two independently designed MOs is negligible (Skromne & Prince 2008).

If the phenotype is caused by general morpholino toxicity then the odds are not negligible. As the majority of non-specific morpholino toxicity in zebrafish is mediated by the p53 pathway (Robu et al. 2007) it is possible to significantly reduce it by either knocking down *p53* with a coinjection or by using *p53* null zebrafish(Gerety & Wilkinson 2011; Bedell et al. 2011; Skromne & Prince 2008). This is why both smoc1tMO and smoc1scMO were co-injected with p53 MO. As these co-injections produced the same range of phenotypes observed in the preliminary work **(Fig. 3.2.1 A and B;** David Sexton 2010**),** this demonstrates that

the C2-C4 phenotype is not dependent on the p53 pathway. p53 associated toxicity can therefore be ruled out.

The best evidence that a morpholino is specific (with the exception of phenocopying targeted null mutations) is that it can be rescued by artificially reintroducing the depleted protein and managing to fix the genetic abnormality. This usually takes the form of injecting artificially transcribed mRNA. If the morpholino is splice targeting then it will not work on a artificial mRNA which excludes introns. If the morpholino is translation blocking then either the target site needs to be re-engineered or non-homologous RNA is used. Though highly desirable, rescue is often unobtainable (Skromne & Prince 2008). This is due to the fact that rescue often results in a gain of function. Injected mRNA is often ubiquitously expressed and it is almost impossible to recapitulate the precise dosage of the introduced gene, which means that it is difficult to completely rescue morpholino knockdown if the gene's function is particularly sensitive to where it is localised or its concentration(Skromne & Prince 2008).

Initially this was the strategy which was pursued with *zsmoc1,* which was unsuccessful **(Fig. 3.2.5)**. However, the *zsmoc1* transcript was based on the sequence recorded in the UNSC genome browser, which excludes both exons 14 and 15 **(Fig. 2.2.1)**. While these exons only account for 15 amino acids, their absence provides a plausible reason for both the failure of the rescue and the absence of an overexpression phenotype **(Fig. 3.2.7 A)**, despite the fact that *hSMOC1* overexpression has a very dramatic phenotype. However, it was possible

to partially rescue the expression of three of the four dorsal ventral marker genes by injecting *hSMOC1* mRNA **(Fig. 3.2.7)**, which indicates that the down-regulation of these dorsoventral genes is specifically due to the loss of smoc-1.

More recently the availability of more and more efficient DNA engineering technologies in zebrafish has allowed the generation of unprecedented numbers of targeted knockouts in zebrafish which has thrown into question the primacy of mRNA rescue for ruling out off target effects (Blackburn et al. 2013; Chang et al. 2013; Cade et al. 2012; Schulte-Merker & Stainier 2014). There are reports that only the phenotypes of 1 in 10 of the morphants reported in the literature phenocopying their associated DNA null, even for those which reported rescue (van Impel et al. 2014; Schulte-Merker & Stainier 2014). Some of this is probably due to the fact that morpholino knockdown is not equivalent to a genetic knockout. For example translation blocking morpholinos will ablate maternally deposited mRNAs while genetic ablation will not. Others may be due to a combination of toxicology and the effect of targeted knockdowns.

In some ways this is being addressed by our use of marker genes to introduce a quantitative rather than a qualitative measure of rescue**(Fig. 3.2.2; Fig. 3.2.3; Fig. 3.2.4; Fig. 3.2.5; Fig. 3.2.6 and Fig. 3.2.7)**. However to further verify the results, efforts were made to utilise TALEN technology to generate *zsmoc1* nulls zebrafish lines in order to confirm whether they mirrored our results. This work is detailed in **chapter 4**.

The live imaging assays of BRE:gfp revealed that *zsmoc1* depletion led to significant increase in bmp signalling intensity in the zebrafish eye, which is consistent with the predictions that smoc-1 is an antagonist of the bmp signalling localised to the ventral retina. While there were no gross defects in eye development caused by MO treatment, these wouldn't necessarily be predicted in zebrafish. The dorsal ventral bmp gradient of the zebrafish eye is known to be required for correct specification of the dorsal retina. When this gradient is ablated by knockdown of *gdf6a* with gdf6aMO it results in a variety of eye malformations ranging from anophthalmia through to coloboma and lens extrusion (Asai-Coakwell et al. 2007). This is likely due to redundancy; the bmp ligands bmp2b and bmp4 are both known to be expressed within the eye along with gdf6a. Of the two, bmp2b is the more significant as the optic cups found in dorsalised *bmp2b* morphants are completely ventralised. While *bmp2b* morphants are grossly dorsalised (C4-C5), they do initially develop normal eye fields which become apparent through sectioning. It is worth noting that both gdf6a and bmp2b are involved in earlier stages of eye development and these ocular abnormalities may be down to the destruction of dorsal ventral patterning during epiboly. When bmp expression was inhibited within the developing eye by the ectopic expression of *smad7* (using the *rx3* promoter), the eyes were largely identical by the ten somite stage, though the expression of the dorsal retinal marker *tbx5* was down regulated(French et al. 2009).

While ablation of the dorsal-ventral bmp gradient does lead to large observable defects in eye formation, the reverse is not true. When the dorsal

identity is up regulated, the retina is dorsalised but microphthalmia/anophthalmia or coloboma defects are not induced. A good example of this is coexpression of *gdf6a* and *bmpr1* within the eye, which leads to the expansion of dorsal marker genes but no large morphological defects (French et al. 2009).

*zsmoc1* morphants do have increased bmp signalling activity. Based on the known examples of up regulated bmp signalling in the eye, this should result in a dorsalised retina. Although this remains to be confirmed.

## 3.3.4 Concluding remarks

Designing and validating a novel qualitative assay of dorsalisation has enabled the confirmation of whether the *zsmoc1* morphant phenotype is truly dorsalised. Based on the results obtained in (**Fig. 3.2.4)**, it is possible to conclude that the embryos are not classically dorsalised by smoc1 depletion. The fact that it is possible to partially rescue *zsmoc1* morphants using *hSMOC1* mRNA, demonstrates that the disruption to the dorsal ventral marker genes is attributable to *zsmoc1* depletion **(Fig. 3.2.8)**. However whole mount immunohistochemistry revealed smad signalling was significantly disrupted during gastrulation(Fig 2.2.16). While the dorsoventral gradient was not affected by smoc1 depletion the mesolateral gradient was disrupted. The gradient was extended dorsoventrally. This disruption could plausibly inhibit the mesolateral migration of the endoderm disrupting epiboly. This would result in defects during neurulation and mesolateral thickening and apparent dorsalised phenotype observance in Fig 3.2.1 but without the corresponding up-regulation of dorsal markers and down-regulation of ventral markers. This is supported by the phenotype of embryos injected with *hSMOC1* mRNA which showed both classical dorsalisation and a corresponding upregulation of dorsal marker genes and down-regulation of ventral marker genes. Unfortunately given the scope of our results it's impossible to tell whether zsmoc-1 actively controls bmp signalling during neurulation (outside of the optic cup) as *zsmoc1* morphants first show defects during

gastrulation, and defects in the establishment of an effective body plan could plausibly account for the defects seen during neurulation and somitogenesis,

Rather than acting as a simple bmp antagonist, the results all support the principle that zsmoc-1 action is conserved with *Drosophila* Pentagone, whereby zsmoc-1 controls the shape of the bmp gradient by acting in a feedback loop with bmp signalling.

While I was unable to replicate the eye defects observed by Abouzeid et al. (2011)**(Fig 3.2.15)**, my results did show that BRE activity is significantly reduced in the embryonic optic cup of *zsmoc1* morphants. Which suggests that bmp signaling is inhibited by zsmoc1 in the optic cup. If this is the case then the retina of the zebrafish should be dorsalised in zebrafish morphants. Further work needs to be done to confirm this.

# Chapter 4

# Establishing and streamlining the production of TALE nucleases

# 4.1 Introduction

## 4.1.1 Traditional genomic editing in zebrafish

While data on SMOC-1 function has been obtained using morpholino knockdown of *zsmoc1* there is always the concern that the phenotype is due to non-specific off target effects and toxicity (Bedell et al. 2011;Gerety & Wilkinson 2011). Ideally, the cleanest genomic data is achieved via induced mutation of the genes in question. This has historically been difficult in zebrafish (Ekker 2008). Traditionally it was achieved in two ways. The first and oldest is high throughput mutation analysis (tilling) following chemical mutagenesis. Tilling can identify many mutations but it relies on random chance to hit the gene of interest. So it is not possible to choose the desired location of your mutations and there is also the significant problem of getting second hits which will cause off target effects. It is also expensive and time-consuming (Huang et al. 2012). More recently, the invention of zinc finger endonucleases (ZFNs) allowed the generation of targeted mutations in zebrafish. First discovered in 1996, ZFNs are endonucleases (typically the cleavage domain of the restriction enzyme Fok1) bound to a zinc finger DNA binding domain, which allows the endonucleases to cleave the target site DNA and induce mutations via non-homologous end joining (Kim et al. 1996). ZFNs have been used to induce targeted mutations in model organisms since 2003 (Bibikova

et al. 2003) and the first zebrafish knockout was produced in 2008 (Ekker 2008). However, targeting zinc finger binding domains is not a trivial task. There is no simple zinc finger targeting code and the DNA binding activity is highly context dependent. Ramirez et al. (2008) showed that even using previously published zinc finger DNA binding modules, 79% of the zinc fingers they synthesised using context-dependent assembly (CoDA) failed to bind to target DNA. This demonstrates that even identical zinc finger arrays known to bind a specific sequence can fail to bind to exactly the same sequence when it is located in a different genomic region. Since ZFNs have to bind in pairs their failure rate would be even higher (Ramirez et al. 2008). Targeting is also highly restricted, depending on the synthesis method. Compatible target sites are on average found every 140-400bp (Blackburn et al. 2013) so there are often genes which have no suitable target sites within them (Clark et al. 2011). In addition ZFNs are proprietary technology and therefore expensive.

Due to the above limitations of ZFN technology, it was decided that a TALEN based strategy would be pursued for the production of a *zsmoc1* null zebrafish line.

## 4.1.2 TALEN proteins

Recently, a DNA binding protein has been characterised that binds to DNA with a simple repeat code. Transcription Activator-Like Effectors (TALEs) are proteins that are secreted by *Xanthomonas* bacteria into their plant hosts to control gene expression (Moscou & Bogdanove 2009). They have a highly repetitive DNA binding domain made up of multiple repeat subunits called residual variable dimers (RVDs; Bogdanove et al. 2010**; Fig. 4.1.1 A)**. These are 33 amino acid long sequences that are identical apart from two amino acids, the 12th and 13th. Crystallography has demonstrated that these two amino acids impart binding specificity to the RVD (Deng et al. 2012; Mak et al. 2012). It is the 13th residue that directly binds the base via the major groove and the 12th residue stabilises the RVD structure. Subunits have now been identified which will bind to each base in the genomic code (NI to A, NG to T, HD to C and NH to G) (Moscou & Bogdanove 2009; Cong et al. 2012; Boch et al. 2009; Miller et al. 2011). By stringing these subunits together, it should theoretically be possible to target any conceivable arrangement of base pairs.

As with ZFNs, (Kim et al. 1996) it has been demonstrated that you can fuse a TALE protein with the catalytic domain of the endonuclease (FokI) and maintain cutting activity **(Fig. 4.1.1 C,D).** These TALEN constructs have been used to generate double stranded breaks in vitro and resultant mutation in vivo (Sander, et al. 2011; Cade et al. 2012; Cermak et al. 2011). In reality there are other targeting constraints, not all of which have been identified **(Ch4.2.1**; Streubel et al. 2012).

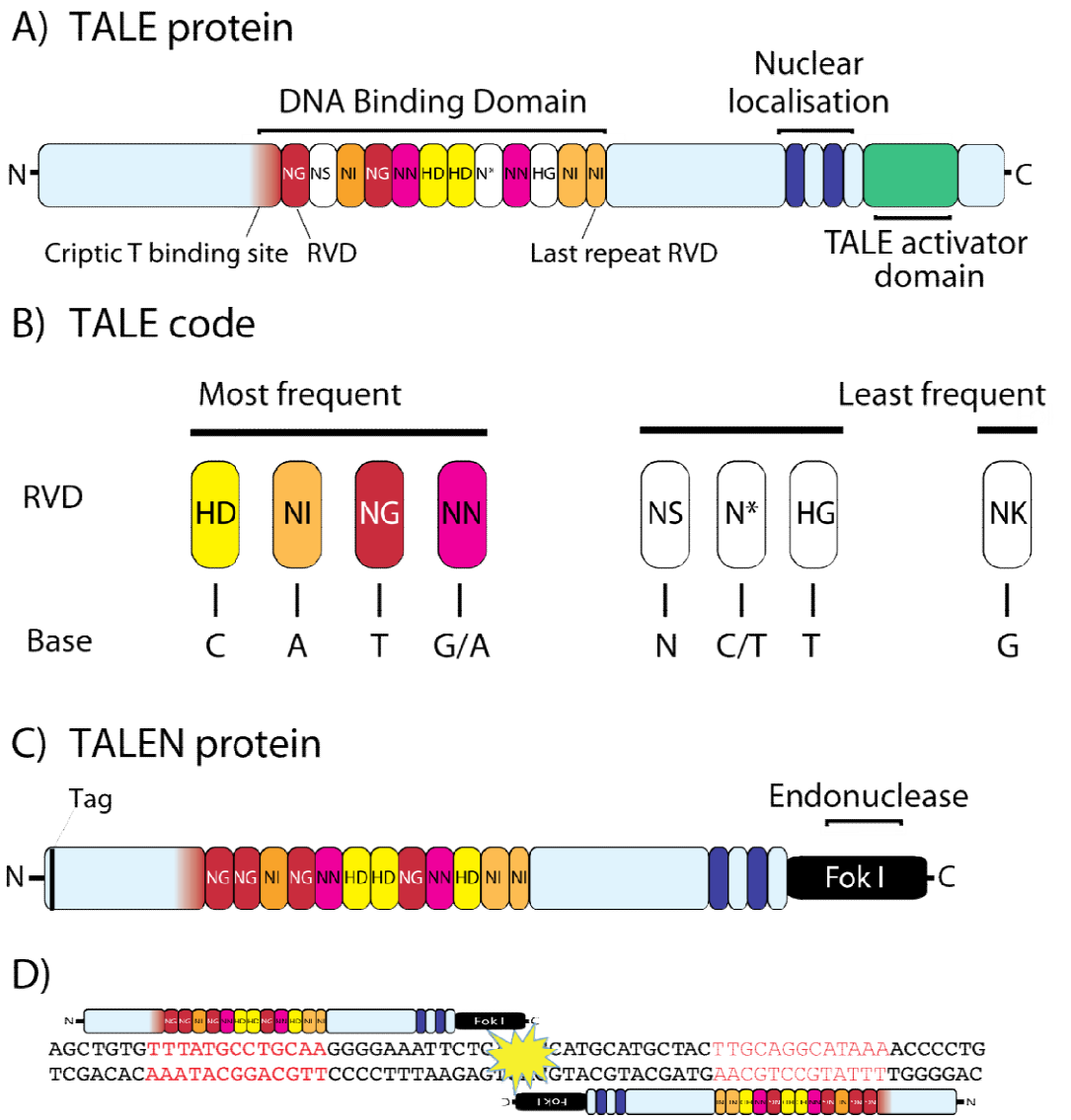



**Fig. 4.1.1 TALE and TALEN proteins, their structure and action.**

**(A)** A generic TALE, it details the main features of a TALE protein, including nuclear localisation sequences and TALE activator domain. The DNA binding domain is composed of between 8-30 residual variable dimers (RVDs) which are 32-33 amino acid repeats. **(B)** The diagram illustrates the TALE code. Diagram shows the seven most common RVD subunits (each is named after the variation in the 12$^{th}$ and 13$^{th}$ amino acid of the repeat positions, * indicates a missing residue) along with the corresponding base they recognise. Arranged from most to least frequent in nature (Moscou & Bogdanove 2009; Boch et al. 2009) **(C)** A TALEN protein: The DNA binding domain and nuclear localisation sequences have been retained from the TALE protein but the protein has been truncated and both a FokI catalytic domain and a protein tag have been added.**(D)** Illustration that of TALEN pair inducing double strand breaks in DNA. The diagram shows that both TALEN proteins must be close enough together for the FokI subunits to dimerise before endonuclease activity is induced (Kim & Kim 2014).

It is possible to fuse the entire FokI DNA cleavage domain with a single TALE backbone (Kim et al. 1996). This is rarely done as use of ZFNs showed that by separating the catalytic domain between two zinc finger backbones it was possible to both increase cleavage efficiency and dramatically reduce off target cutting and toxicity (Mani et al. 2005). This is because it increases on site specificity as both proteins must bind to induce DSB. Most TALENs are engineered as pairs which flank a short spacer region. When they bind to the target sequence they are in opposing orientations which brings the FokI domains into close proximity inducing DNA cleavage **(Fig. 4.1.1 D)**.

The initial studies show great potential. TALENs have far fewer binding constraints than ZFNs with the average gene containing hundreds of potential target sites (on average a target site is present every 1-3bp;Kim & Kim 2014) . The reports of TALEN activity in zebrafish are promising. TALENs have been shown to have high cleavage efficiencies of between 20%–76.8% (Moore et al. 2012;Dahlem et al. 2012) when compared to ZFNs cleavage efficiencies of 1.1%–3.3%. While Chen et al. showed that on average a given TALEN pair produces 10 times the deletions of equivalent ZFN construct in zebrafish (Chen et al. 2013) with comparable or lower toxicity (Moore et al. 2012; Dahlem et al. 2012). They have been shown to have fewer off target effects than ZFNs, probably due to greater specificity. A good example of this is when Dahlem et al. targeted *ryr1a-ex6* with a TALEN pair which showed 100% cleaved efficiency. The *ryr1* target site showed no detectable cleavage despite the fact that the target sites only differed by 2bp (Dahlem et al. 2012). A point further reinforced by Zhang et al., who demonstrated

that a single nucleotide miss match can reduce cleavage efficiency from 70-50% to just 2% (Zhang et al. 2011).

Coupled with the fact that they are open source, and so considerably cheaper to produce makes them a far more attractive option than ZFNs.

## 4.1.3 TALEN synthesis strategies

Given that the average TALE binding domain used is between 10 and 20 RVDs in length they are highly repetitive (Cong et al. 2012). Virtually all the amino acids in RVD's are identical, and even with codon degeneracy they are extremely repetitive DNA sequences. However artificially produced TALENs are even more repetitive. The reason is twofold, firstly of the known naturally occurring RVDs only five are generally used for gene targeting. The HD, NG and NI were clearly the most suitable for gene targeting, as they are the most abundant in nature, have the strongest binding activity in vivo and show the greatest specificity of any RVD to a specific base (C, A and T respectively) **(Fig. 4.1.1 B;** Cong et al. 2012; Moscou & Bogdanove 2009; Boch et al. 2009). It is this last property which is the most important as this means that these RVDs can be arranged in a particular order to perfectly correspond to the complementary base.

However there is no common natural RVD that has a high specificity for G bases. NN is common in nature, but does not show specific binding preferences between G or A bases. NK has a much greater specificity to G bases but is very rare in nature and has been shown to have quite a low binding strength (Christian et al.

2012; Meckler et al. 2013). The problem may have been solved, as more recently the artificial TALE RVD NH has been engineered which will bind to Guanine and shows a much greater binding affinity to DNA than the NK subunit (Miller et al. 2011; Cong et al. 2012). However, the binding affinity is still not as strong as the NN RVD and TALENs using the NN subunit show greater cleavage efficiency (Meckler et al. 2013).

The repetitive nature of the TALE DNA binding domain makes it impossible to synthesise by solid-phase DNA synthesis and challenging to clone. To address this, novel cloning strategies have been developed by different groups to synthesise custom TALE arrays. This is the second reason that TALEN proteins have such highly a repetitive DNA sequence. To simplify the cloning process the same RVD units are cloned multiple times in different positions(Cermak et al. 2011; Sander, Cade et al. 2011; Reyon et al. 2013; Briggs et al. 2012). So for example the sequence of every NI subunit is identical in a TALEN whether it is the 1st, 7th or 18th RVD in the binding domain.

The first strategy developed was Hierarchal ligation. Hierarchical ligation is based on traditional cloning of individual RVDs into a standard array of plasmids **(Fig. 4.1.2 A).** This strategy is labour-intensive and requires a minimum of 15 ligation reactions to make a DNA binding domain 16 RVDs long. The large number of reactions means it easy to make a cloning error which will disrupt synthesis. Which is why in the restriction enzyme and ligation (REAL) method a large library of plasmids are assembled which contains all possible combinations of RVDs for

lengths between one and four RVDs long(Sander, Cade et al. 2011). However this still requires a library of at least 296 plasmids **(Fig. 4.1.2 A).**

Golden Gate cloning has been of great utility in TALEN synthesis. It has been used to streamline the hierarchical ligation strategy in the REAL method (Sander, Cade et al. 2011) where the final ligation reaction can be conducted simultaneously. However, it is most heavily used in the Voytas method (Cermak et al. 2011). Golden Gate cloning relies on the use of Type II restriction enzymes. These enzymes have a specific binding site (usually a non-palindromic six base pair sequence) but cut upstream of their binding site to produce five prime over hangs which are four base pairs long (Engler et al. 2008). This permits the generation of any desired sequence for the overhang in a plasmid simply by directing site-directed mutagenesis next to the enzymes' binding site. The advantage of using this method is that it can produce ligation of multiple fragments without DNA scars. If the overhangs are too similar mis-ligation can occur. To avoid this only non palandromic overhangs are used for complex assemblies (Engler et al. 2009). Therefore with a type II restriction enzyme which produces only four base pair over hangs, it is possible to specify 12 different sites. This allows you to mix multiple plasmids into a single digest with a single restriction enzyme and ligate DNA from up to 11 plasmids.

Of course, each plasmid could also re-ligate to its parent plasmid as the original cut site will have complementary overhangs. Golden Gate Technology uses two strategies to overcome this problem. The first is to run the reaction through multiple cycles of ligation and digestion, typically between 10 and 20 for ligations

involving over 2 fragments. This is because any reverse ligations will result in a plasmid which can be redigested. As the reaction continues the proportion of original plasmids will decrease as the population of correctly ligated plasmids increases(Cermak et al. 2011; Sander, Cade et al. 2011). The second strategy is selection. Most of the undesired plasmids can be screened out by antibiotic selection. However the destination plasmid must be screened out by an additional method**(Fig. 4.1.2 C)**. Typically this is either with blue/white *lacZ* screening or *ccbd* survival screening(Cermak et al. 2011; Sander, Cade et al. 2011).DFNA

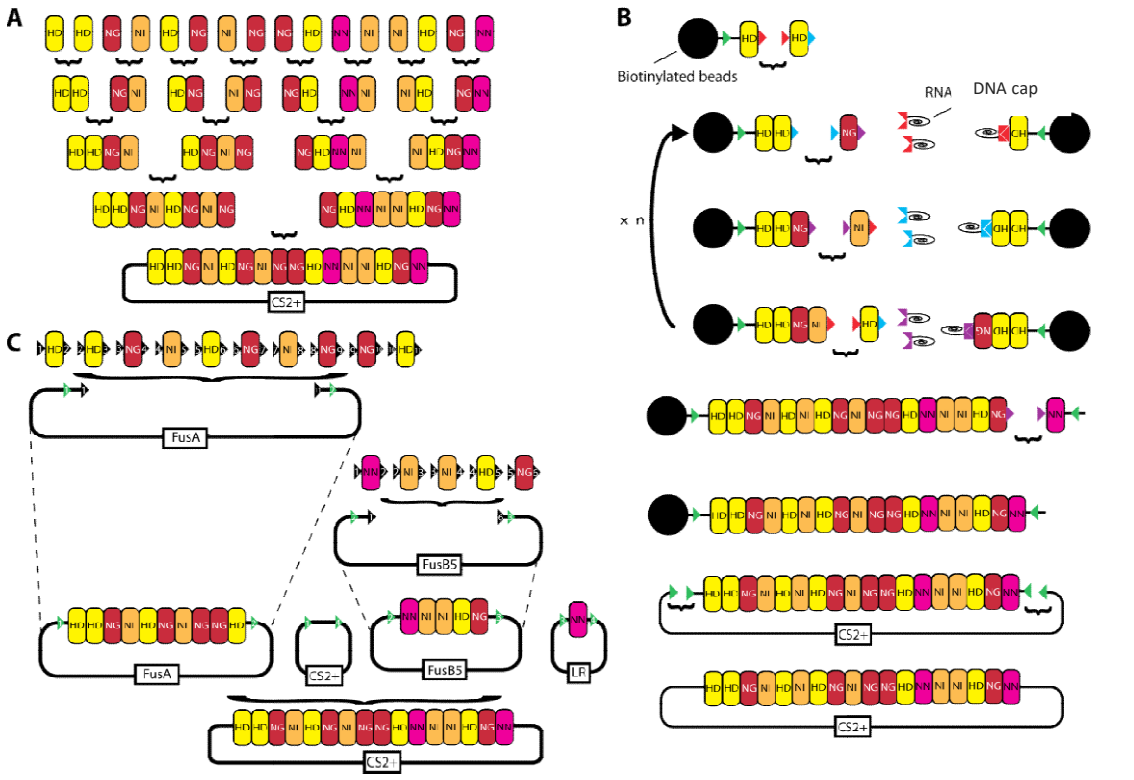



**Fig. 4.1.2 The three main approaches for TALEN synthesis**

**(A)**Hierarchal ligation: traditional cloning is used to build up unique scaffolds used in the REAL synthesis of TALENs. **(B)** Solid phase ligation: ICA **(C)** Golden Gate cloning: As used in the Voytas method.

(Cermak et al. 2011; Sander, Cade et al. 2011; Reyon et al. 2013; Briggs et al. 2012)

The Voytas method uses single RVD plasmids, so it takes three Golden Gate reactions to synthesise a TALEN with a DNA binding domain of over 11 RVDs (given that there are only 12 possible none palindromic overhangs). However by using tetramers and trimers Golden Gate reactions can generate full-length TALENs in a single reaction(Cermak et al. 2011).

While the Golden Gate cloning is attractive for small-scale use, it is a labour-intensive method for producing large numbers of TALENs which has lead to the development of several high throughput methods. These methods rely on solid-phase reactions. The initial sequence of DNA is immobilised on either a streptavin bead (REAL Fast Ligation-based assemble solid-phase high-throughput (FLASH; Reyon et al. 2012; Reyon et al. 2013) or a DNA chip (Iterative Capped Assembly (ICA; Briggs et al. 2012) as the initial seed of ligation, and the desired RVDs are ligated in a linear sequence**(Fig. 4.1.2 C)**. This arrangement leads to easy automation; TALEN proteins can be assembled in one go without the need for intermediate purification or analysis. As with Golden Gate reactions there is a cloning limit to the number of unique non- palandromic overhangs, to sequence a TALEN in one reaction FLASH uses a large library of plasmids (800<) which allows the ligation of RVD blocks up to 4 subunits long. The large library of plasmids increases the time and cost of optimisation but once the system is up and running the platform can produce 96 TALENs in 24 hours for less than $100 (Reyon et al. 2012). ICA uses a more elegant solution: RVDs are ligated individually but unique none palandromic overhangs can be reused as small oligonucleotides can cap incomplete ligation events (Briggs et al. 2012). Hundreds of TALENs can be

synthesised this way in just three days. However the process requires specialised equipment which would be a considerable investment for most labs.

The Voytas method (Cermak et al. 2011) was chosen as it relies purely on the ligation of multiple plasmids in one reaction using Golden Gate cloning. This is an advantage as it avoids PCR and the resulting mutations. Once optimised, TALENs can be rapidly synthesised in just two weeks (Cermak et al. 2011). While these advantages are shared by a solid phase ligation these protocols are considerably more expensive and time-consuming to optimise and unnecessary for our purposes.

# 4.2 Results

## 4.2.1 TALEN design and targeting

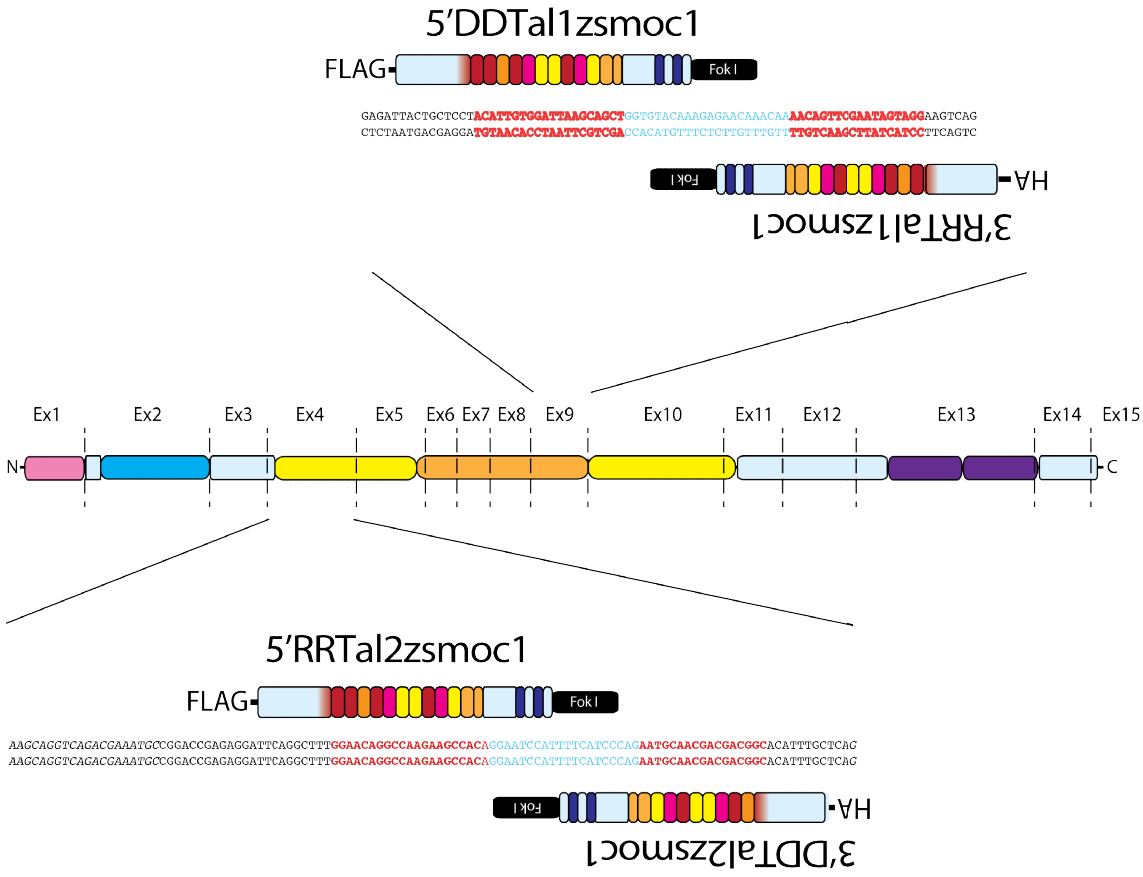



**Fig 4.2.1 The TALEN binding sites.**

TALEN pair TAL1-smoc1 targets exon nine while TALEN pair TAL2-smoc1 targets exon four.

To create a zebrafish knockout of *zsmoc1* two TALEN pairs were designed to target this region. Targets were designed using the TAL effector targeter (v1.0 Cornell University; Doyle et al. 2012). At this time TALE technology was in its

infancy, so it was decided that the TALEN pairs should be designed to mimic the natural structure of the TALE protein (Cermak et al. 2011). Though most of these criteria have now been shown to have little effect on TALEN activity with the exception of preceding the binding site with a thiamine (Doyle et al. 2012). Every TALE DNA binding domain begins with a cryptic RVD which specifies thiamine and all but one natural TALE target site begins with a thiamine (the exception was found by Yu et al. in 2011 and targets a cytosine). While TALENs have been designed which ignore this rule and still induce targeted deletions, the cleavage efficiency generally seems to be greatly reduced (Sun, Ning 2013; Meckler et al. 2013). Another critical factor to TALEN cutting efficiency is the size of the spacer region. Too long and the FokI domains will be separated and won't be able to cleave DNA. Too short and both TALENs will not be able to bind the target site simultaneously. To take this into account we designed TALEN binding sites that would have a spacer 15 - 22bp (Cermak et al. 2011). Even so this leaves an abundance of potential TALEN target sites. To further narrow down our selection the following criteria were used.

**No TALENs were designed against the first two exons.**

Genes often have multiple initiator codons, and given the unusual conservation of the SMOC-1 signal peptide **(Fig. 2.2.2)** it was a real concern that targeted deletions might be corrected by alternative splicing.

**No TALENs were designed against the final exon.**

Any deletions in the final codon would not trigger nonsense mediated decay (NMD) and therefore would be unlikely to give nulls.

**TALENs pairs were designed to be as close to 21 RVDs as possible.**

Longer TALENs have longer target sites so should have greater specificity. This should result in fewer off target cleavages and lower toxicity. Given the limitations of the Voytas method of Golden Gate cloning, target sites greater than 21 bp require additional plasmids. So for utility any pairs longer than this were discarded.

**Target sites were screened against the zebrafish gene using UCSC genome browser to ensure that they were unique.**

Although the chances of finding a random match for a randomly generated sequence of 16bp is theoretically small, genomes are not randomly assembled. By using only unique sequences off target deletions can be avoided.

**The NN rather than NK RVD was used to target Guanine basis.**

NN has been shown to have both higher binding affinity (Meckler et al. 2013) and greater activity than NK (Christian et al. 2012).

**Only TALEN pairs with a high content of both HD and NN subunits spread evenly in the DNA binding domain were picked.**

HD(T) and NN(G-C) have been classified as strong binders. TALEN pairs which have less of these RVDs have shown low cleavage activity. To take this into account TALEN pairs with a high content of HD and NN spread evenly throughout the DNA binding domain, were preferred (Meckler et al. 2013).

**TALEN pairs that targeted strongly conserved sequences in exons four and nine were picked.**

TALEN pairs produce an array of Micro deletions. It was hoped that some would be in frame and produce novel *zsmoc1* hypomorphs in addition to *zsmoc1* nulls. As evolutionarily conserved sequences are likely to be more functionally significant they were targeted over weakly conserved regions.

Multiple TALEN backbones have been generated, though variable efficacies have been reported in the literature about how effective different TALEN backbones are (reported mutation rates range from as low as 1% to as high as 70% (Cade et al. 2012; Dahlem et al. 2012)). I chose to use the paired backbones CS2TAL3RR and CS2TAL3DD (Dahlem et al. 2012). This was for two reasons. Firstly they have been shown to induce mutations at a high frequency in vivo. With golden (melanocyte nulls) knockouts their cleavage efficiency was even high

enough to induce phenotypes in the $F_o$ population. Secondly, these backbones have been engineered to minimize toxicity (Dahlem et al. 2012; Bedell et al. 2012). This has been achieved by using heterodimeric FokI domains instead of a homodimeric FokI. Originally this technique was pioneered in ZFNs (Miller et al. 2007). The FokI catalytic domains were mutated at either Q486E/I499L(EI) or E490K/I538K(KK). This makes it impossible for ZFN proteins bearing the same mutations to dimerise together. However, they can bind to a FokI catalytic domain containing the other mutation. If five prime and three prime targeting ZFNs use the opposing scaffolds (either EI/KK or KK/EI) then the chance of random off target dimerisation is reduced by two thirds (Miller et al. 2007) as homodimers (i.e. EI/EI and KK/KK) are catalytically inactive. Subsequently Doyon et al. (2011) improved on the principle by inducing mutations to produce another pair of heterodimeric FokI domains, Q486E/I499L/N496D (ELD) and E490K/I538K/H537R (KKR). This pair was designed not only to inhibit homodimer formation but also to promote heterodimer binding. Which results in increased DNA cleavage activity. This was demonstrated both in vitro and in human cells (Doyon et al. 2011). It was the Doyon FokI domains utilised in the CS2TAL3RR and CS2TAL3DD scaffolds. As predicted the backbones have been shown to have reduced toxicity (Dahlem et al. 2012; Bedell et al. 2012).

## 4.2.2 TALEN construction

With TALEN construction it was vital that the RVDs were correctly assembled. TALEN pairs in which a single base pair does not match the target site have a five-fold decrease in cleavage efficiency (Zhang et al. 2011). The plasmid kit used was the Voytas V1.0 and was obtained from Peter Hohenstein and the Hastie lab. Before they were used they were assessed by Sanger sequencing to confirm that every plasmid was correctly labeled and free of mutations.

Once assembled each plasmid was screened with a restriction digest to confirm that cleavage had occurred. Plasmids in the first round of Golden Gate cloning were digested with Esp3I which produced fragments of the RVDs. For the second round of digestions it was impossible to find good restriction sites located near the DNA binding domain so it was necessary to excise the entire TALEN construct using NotI and KpnI. The relevant plasmids maps are located in the appendix **(Sup. Fig. 4. and Sup. Fig. 5).**

Each RVD was 99bp long but only contained two amino acids which varied. It was the six base pairs that coded these residues were the critical sequences **(Fig. 4.2.3).** To identify these sequences automatically we targeted the six base pairs upstream. These coded for LQ residues and were unique in the RVDs. The chromatogram of the upstream 6bp of these were manually assessed and the order annotated. As artificial RVDs are highly repetitive sequencing can only be conducted from primers designed to flank the target region (C8F and C8R pFus plasmids and TalF and TalSeq for constructs). Sanger sequencing produced reads of up to 900 bp. Since the pfus constructs are between 750 and 1050bp long and

are sequenced in both the forward and reverse direction, it is possible to sequence them in their entirety. So for the initial TALEN assembly it was possible to ascertain the precise RVD order of the assembled plasmids. For each sequencing reaction three plasmids were picked (unless fewer than 3 colonies were obtained) and independently sequenced.

---

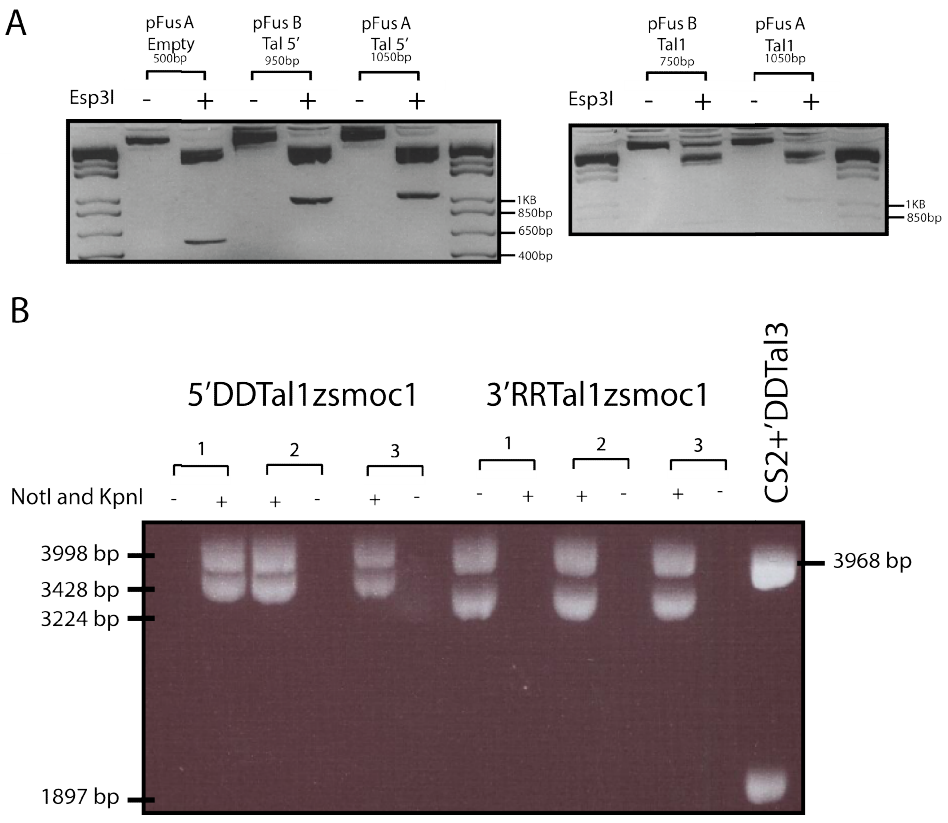



**Fig 4.2.2 Example restriction digestions used to confirm the length of RVD insertion.**

(A) Restriction digests using Exp3I for the pFus plasmids selected for 5'DDTal1zsmoc1(pFusA_2 and pFusB9_2) and 3'RRTal1zsmoc1(pFusA1 and pFusB7_2). **(B)** Digests using Not1 and KpnI for 5'DDTal1zsmoc1(colonies 1, 2 and 3) and 3'RRTal1zsmoc1 (colonies 1, 2 and 3).

A

| Sequence | | Amino Acid | Description |
|---|---|---|---|
| fw | CTCCAG | LQ | Invariant 5' sequence of artificial RVDs |
| rv | CTGGAG | | |
| fw | CCACGA | HD | Unique sequence for the HD subunit |
| rv | TCGTGG | | |
| fw | CAACAT | NI | Unique sequence for the NI subunit |
| rv | ATGTTG | | |
| fw | CAACGG | NG | Unique sequence for the NG subunit |
| rv | CCGTTC | | |
| fw | CAACAA | NN | Unique sequence for the NN subunit |
| rv | TTGTTG | | |

B

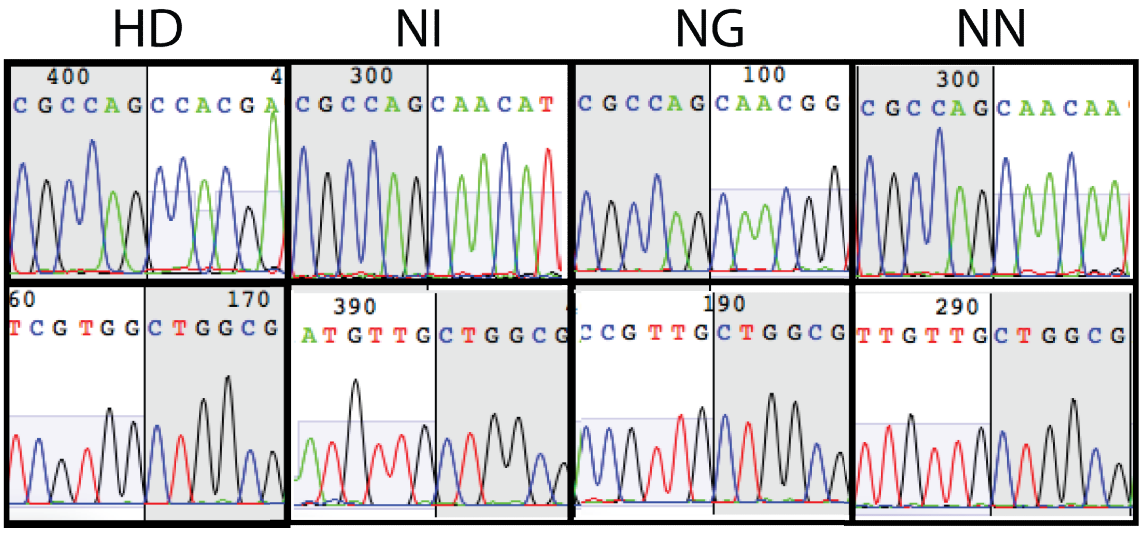



**Fig 4.2.3 Confirmation of the RVD structure after cloning using Sanger sequencing.**

**(A)** Table summarizing the forward and reverse sequences of the two variable amino acids in the four RVDs used in town synthesis along with the invariant sequence of two amino acids (6 bp) upstream of the variable residues. **(B)** Example chromatograms for each of the RVDs used in the synthesis of synthetic TALENs in both the forward and reverse directions. Variable sequence is shown in white while the invariant sequence is highlighted in light blue. Chromatograms were taken in APE v4.0.

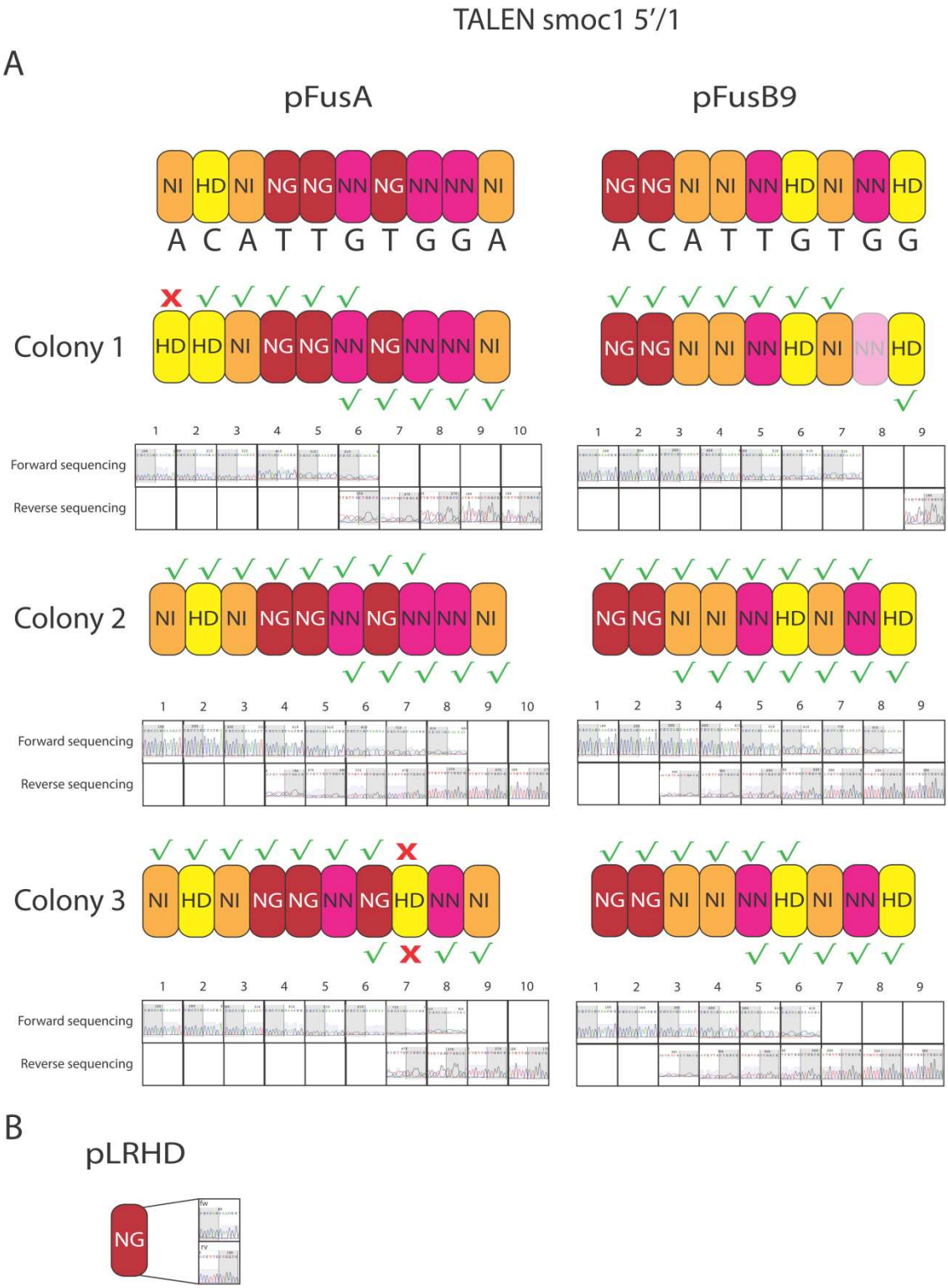



**Fig 4.2.4 The results of the first round of Golden Gate cloning to synthesise 5'Tal1zsmoc1.**

**(A)** In the initial round of the cloning for 5'Tal1zsmoc1 was successful and three colonies for the FusA plasmid and three colonies for the FusB plasmid were selected for sequencing. Of the FusA clones only colony 2 was free of cloning errors. No cloning errors were detected in FusB9 though the sequencing data for colony 1 was incomplete so an error could be present in RVD eight **(B).** Sequencing data for the last repeat plasmid confirming it codes an NG repeat and that it is free of mutations.

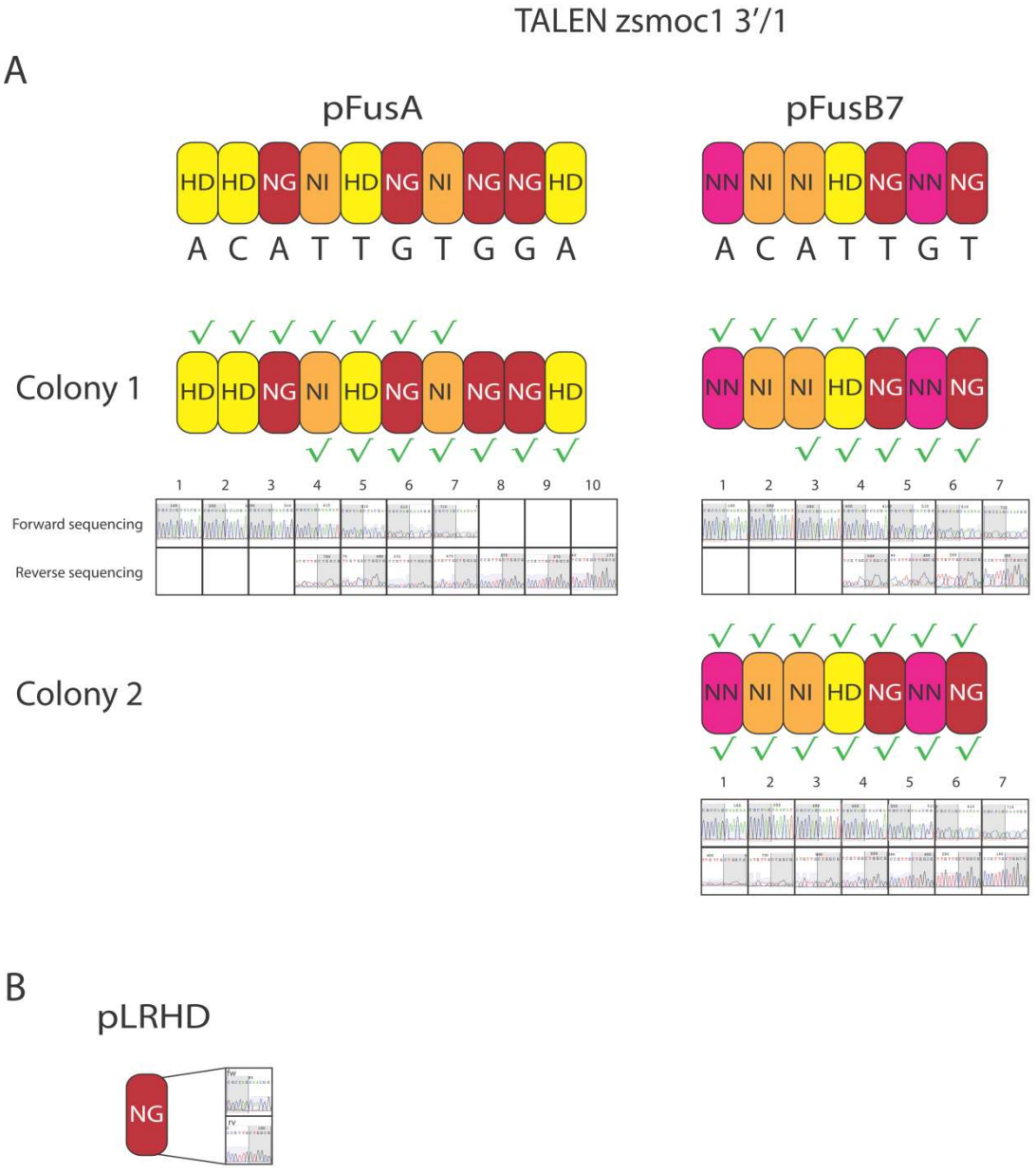



**Fig 4.2.5 The results of the first round of Golden Gate cloning to synthesise 3'Tal1zsmoc1.**

**(A)** In the initial round , the cloning for 3'TALzsmoc1 only one colony was isolated for the FusA plasmid and two colonies for the FusB plasmid were sequenced. No sequencing anomalies were detected. **(B)** Sequencing data for the last repeat plasmid confirming that it coded an NG repeat and it is free of mutations.

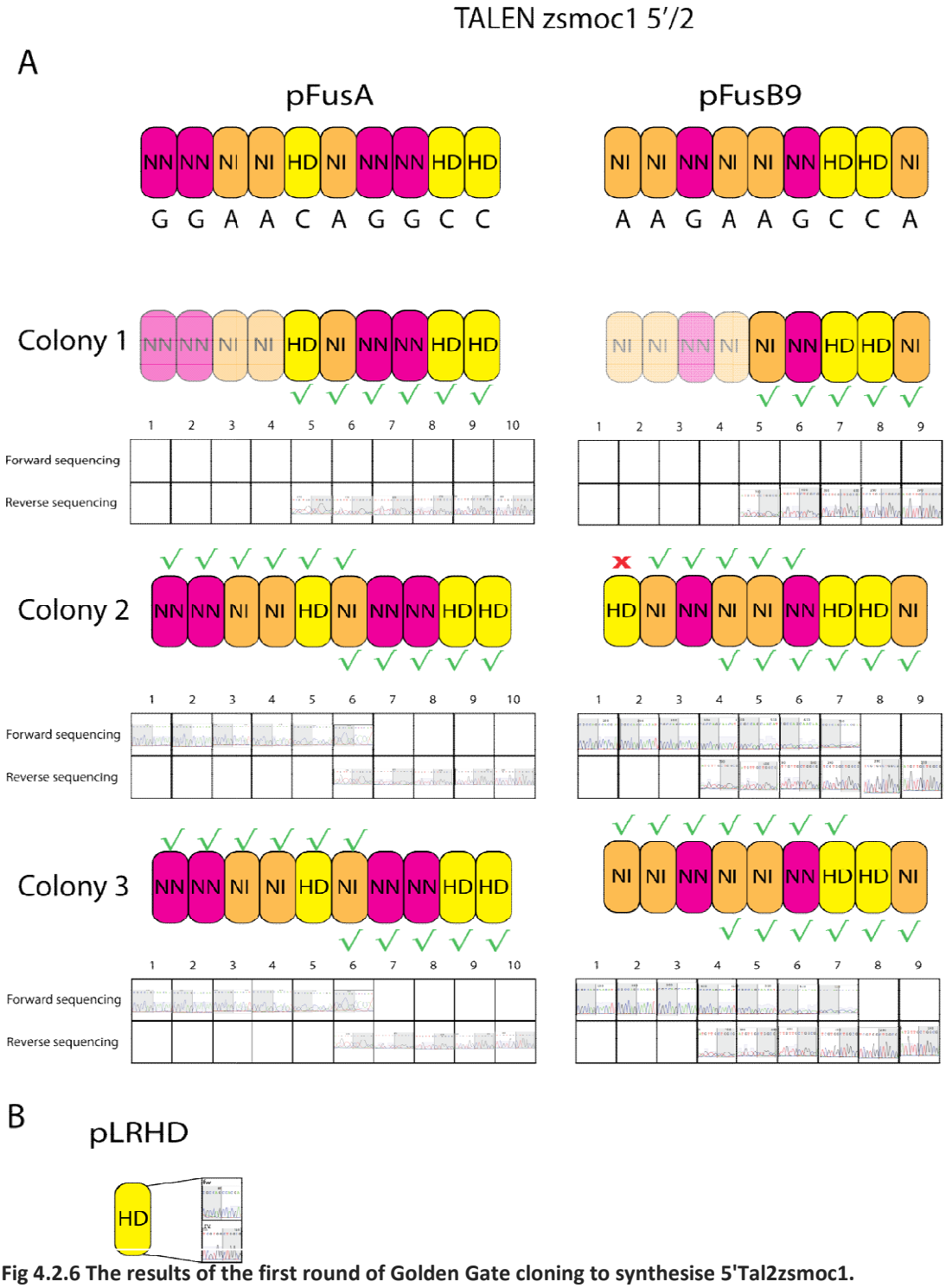



**Fig 4.2.6 The results of the first round of Golden Gate cloning to synthesise 5'Tal2zsmoc1.**

**(A)** In the initial round, cloning of 5'Tal2zsmoc1 was successful and three FusA plasmid and three colonies for the FusB plasmid were selected for sequencing. No cloning errors were detected in FusA Golden Gate reactions though forward sequencing for colony 1 failed. Of the FusB clones only colony three could be fully verified as free of errors. The sequencing of colony one is incomplete and the first RVD in colony two is incorrect. **(B)** Sequencing data for the last repeat plasmid confirming it codes an NG repeat that it is free of mutations.

*RVDs predicted by fragment size but not confirmed by sequencing were shown as transparent boxes.

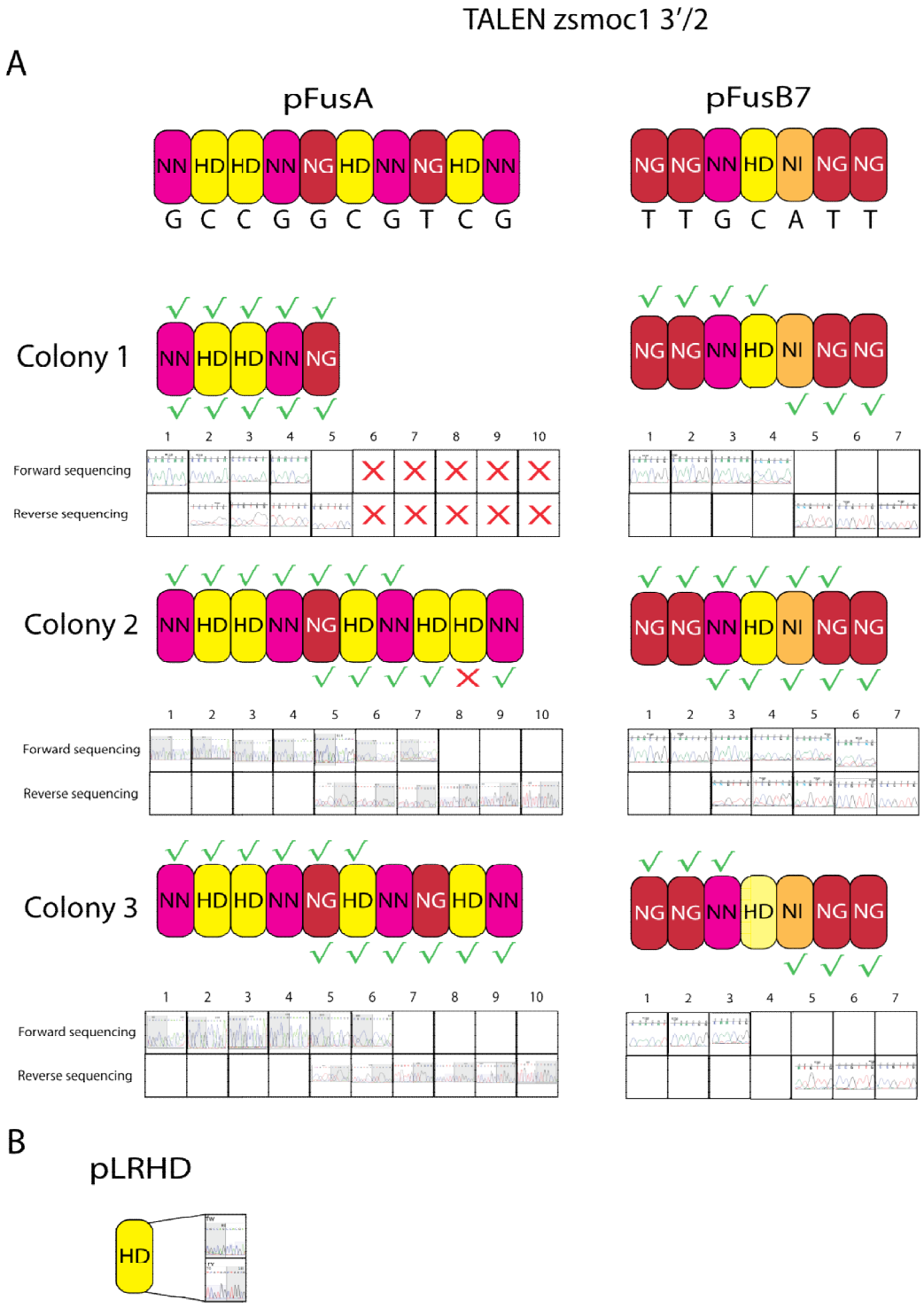



**Fig 4.2.7 The results of the first round of Golden Gate cloning to synthesise 3'TAL2zsmoc1.**

**(A)** In the initial round, cloning of 5'TAL2zsmoc1 was successful and three colonies for both FusA and FusB plasmid were selected for sequencing. Only the third colony isolated for FusA has been correctly assembled. The Golden Gate reaction that produced Colony one has failed resulting in a truncated product. The second last RVD in colony 2 is incorrect. Of the FusB clones both colonies one and two could be fully verified as free of errors. The sequencing for colony 4 is in complete and the first RVD in colony two is incorrect. **(B)** Sequencing data for the last repeat plasmid confirming that it codes an NG repeat that is free of mutations.

Those constructs which were correctly assembled and completely sequenced were selected for the next round of Golden Gate reactions. In the case that more than one construct was fully sequenced and free of errors the colony which had the greatest sequencing coverage was picked. So the first TALEN pair was made up of 5' pFusA_2 pFusB9_2 **(Fig. 4.2.4)** and 3' pFusA_1 pFusB9_2. **(Fig. 4.2.5)** While from the second synthesis plasmids pFusA_3 pFusB9_3 were picked for 5' construct **(Fig. 4.2.6)** and plasmids pFusA_3 pFusB9_2 were picked for 3' construct **(Fig. 4.2.7).** The last repeat (LR) plasmid was also sequenced for both constructs, ensuring they coded for the correct RVD.

Once the full TALEN binding domain is synthesised it is between 1800 and 2000 bp. Given that each sequencing run was 800 bp at most, it is impossible to fully validate the entire binding domain using Sanger sequencing. However, the second Golden Gate reaction contains only three plasmids. Therefore it is possible to infer the success of the cloning reaction as long as the sequencing read covers the 5' end of pFusA and the reverse read crosses the LR RVD into the 3' end of pFusB. If the restriction digestion proves that the binding domain of the TALEN construct is the correct size **(Fig. 4.2.3 B)** then the whole fragment is almost certainly intact. As the Voytas method is entirely ligation based mutations are very unlikely to have occurred (Cermak et al. 2011).

The second round of Golden Gate ligations were successful. The number of colonies produced was greater (Data not shown) and every colony sequenced was correctly assembled **(Fig. 4.2.8; Fig. 4.2.9).**

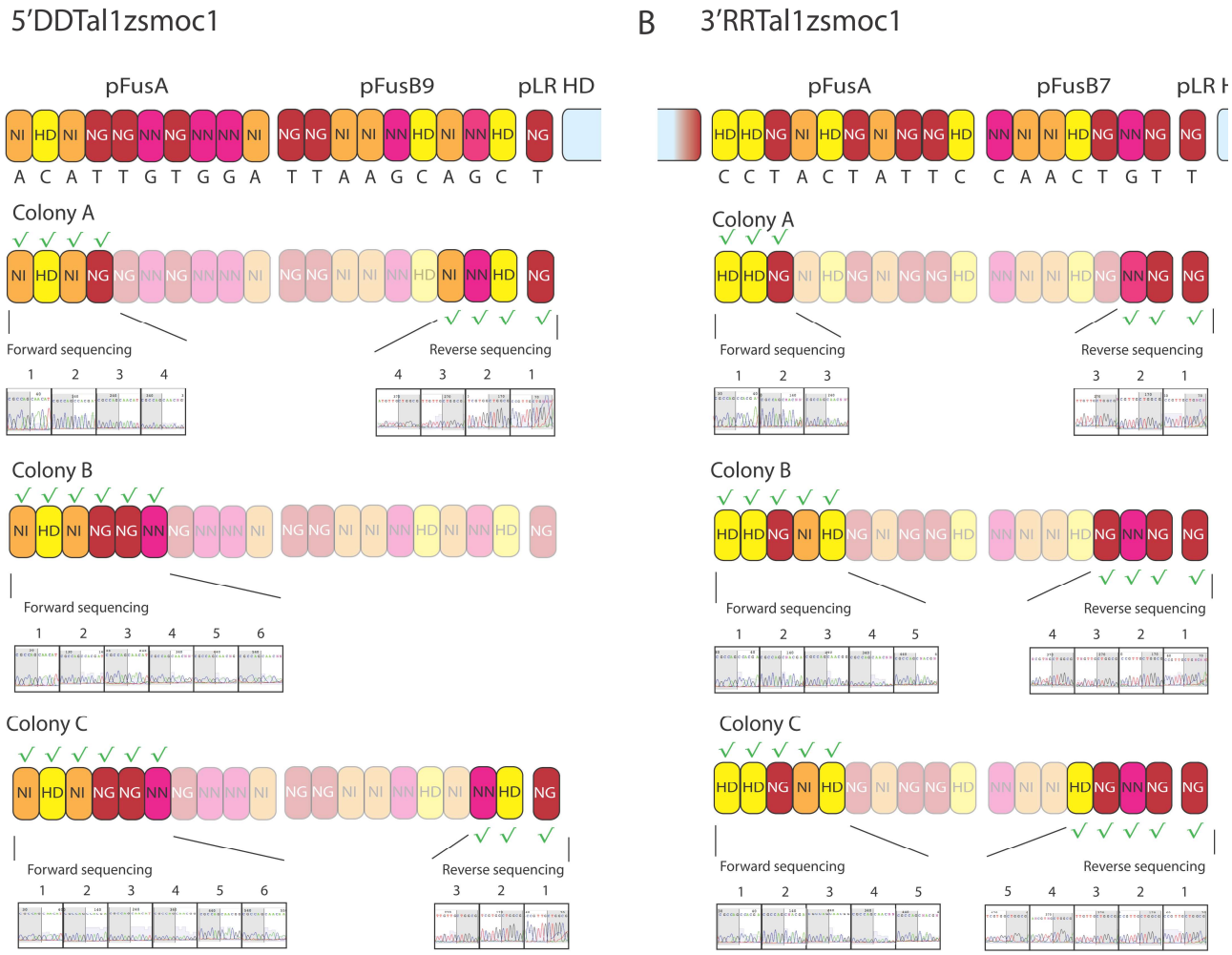



**Fig. 4.2.8 The second Golden Gate cloning reaction for the synthesis of 5'DD and 3'RRTAL1zsmoc1**

For both 5'DDTAL1zsmoc1 **(A)** and 3'RRTAL1zsmoc1 **(B)** three colonies were picked for sequencing. Due to the highly repetitive sequence of the artificial TALEN DNA binding domain it was impossible to sequence the full-length of the TALE binding domain. RVDs confirmed by sequencing are denoted by Green ticks and shown in full colour above the corresponding sequencing traces. Those RVDs merely inferred by total fragment length (confirmed by restriction digests) are depicted by transparent RVD boxes.
Every colony sequenced was consistent with the predicted structure.

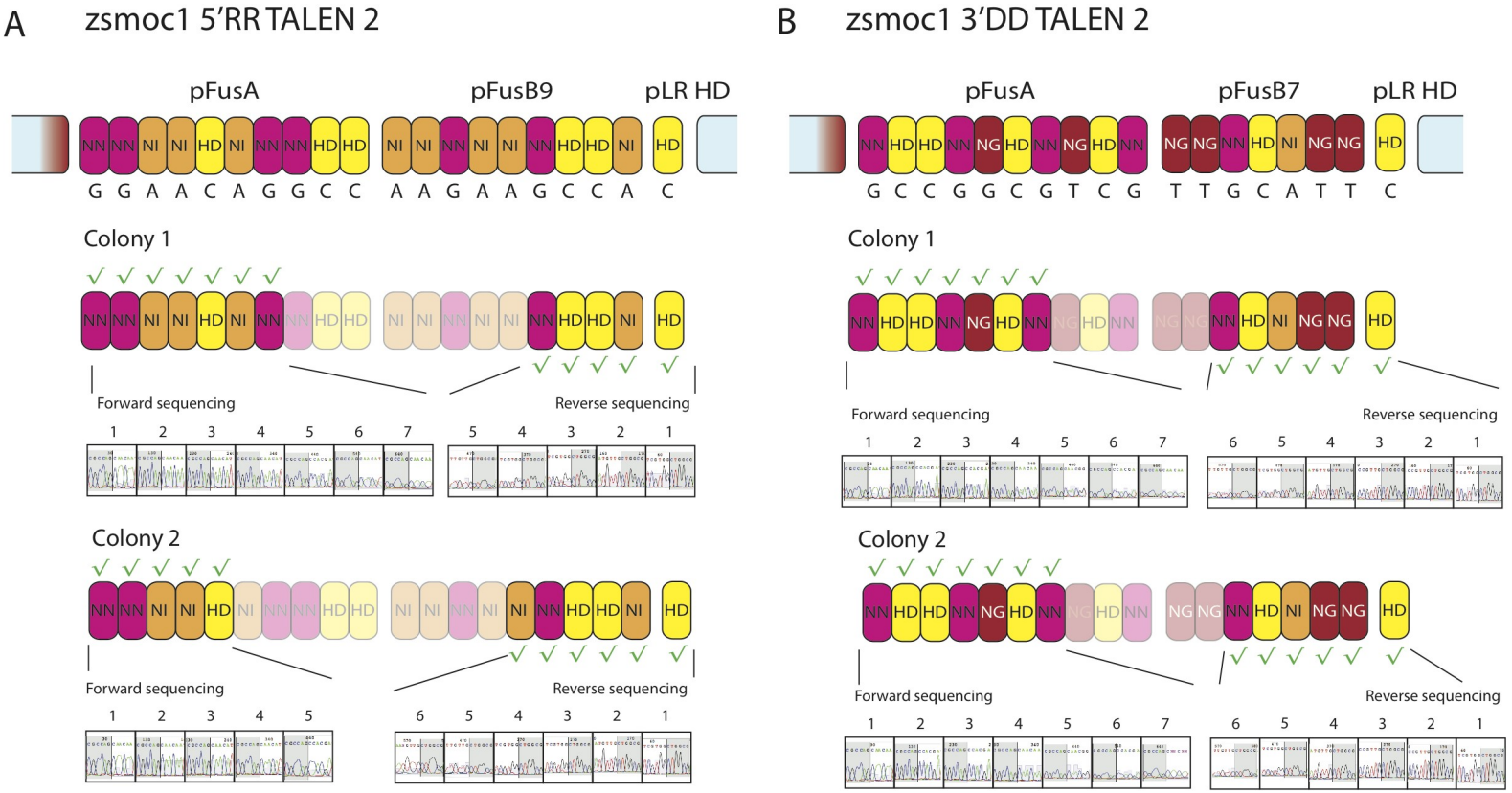



**Fig. 4.2.9 The second Golden Gate cloning reaction for the synthesis of 5'RR and 3'DDTAL2zsmoc1**

For both 5'RRTAL2zsmoc1**(A)** and 3'DDTAL2zsmoc1**(B)** two colonies were picked for sequencing. Due to the highly repetitive sequence of the artificial TALEN DNA binding domain it was impossible to sequence the full-length of the TALE binding domain. RVDs confirmed by sequencing are denoted by Green ticks and shown in full colour above the corresponding sequencing traces while those RVDs merely inferred by total fragment length (confirmed by restriction digests) are depicted by transparent RVD boxes.
Every colony sequenced was consistent with the predicted structure.

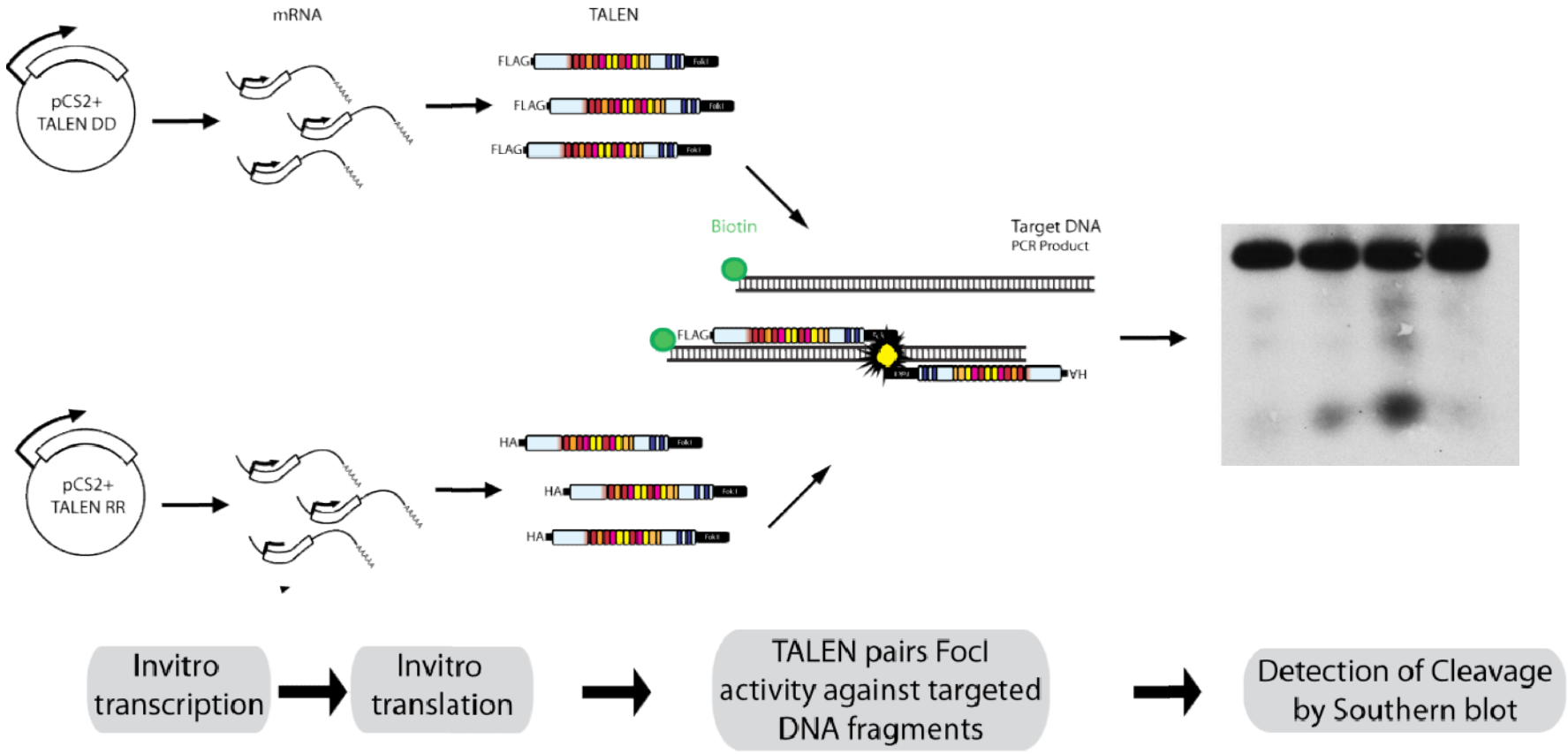
## 4.2.3 In vitro confirmation of TALEN endonuclease activity

To rapidly determine the cleavage efficiency of our TALEN constructs a novel in vitro assay was developed **(Fig. 4.2.10).** Both constructs of the TALEN pairs were transcribed using the mMessage mMachine Kits (Ambion) and the RNA was translated using the Retic Lysate IVT™ Kit (Ambion). All successfully translated proteins were then mixed along with biotinylated DNA. The DNA has either previously been isolated from the target exon or an unrelated exon by PCR amplification with biotinylated primers.

In a correctly assembled and active TALEN pair both constructs should bind to their target sites on the target DNA, bringing their FokI domains close enough together to induce double-stranded breaks. The presence of these lighter cleaved DNA fragments is assessed using a southern blot.

**Fig. 4.2.10 Schematic of *in vitro* assessment of TALEN pair cleavage strategy using southern blotting.**

mRNA was initially transcribed using the mMessage mMachine kits (Ambion), transcripts were tailed with the Poly(A) Tailing Kit (Ambion) before translation with the Retic Lysate IVT™ Kit (Ambion). TALENs pairs are then mixed and incubated overnight at 28°C with 1 µg of target biotinylated DNA. Half of each sample was then run in a 3% agarose gel. DNA cleavage was then assessed using a southern blot. The membrane was probed with anti-biotin and any biotinylated DNA was detected. If cleavage has occurred a second band will be detected from the cleaved five prime end of the target DNA.



As TALEN proteins encoded by both pCS2Tal3DD and pCS2Tal3RR are tagged (by FLAG and HA tags respectively) it was possible to confirm the translation of both TALEN pairs by sodium dodecyl Polyacrylamide gel electrophoresis sulfate (SDS PAGE) followed by Western blotting. After the retic lysate reaction had been separated by gel electrophoresis and transferred by blotting to PVDF membrane the membrane was probed with anti-HA. This showed that both 3'RRTal1zsmoc1 **(Fig. 4.2.11 A)** and 5'RRTal2zsmoc1 **(Fig. 4.2.11 B)** were both translated to produce single bands. Both of which migrated at a rate consistent with their predicted MW (3'RRTal1zsmoc1 103.19kDa, and 5'RRTal1zsmoc1 118.19 kDa respectively). The membranes were then stripped and re-probed with anti-FLAG. Both 5'DDTal1zsmoc1 **(Fig. 4.2.11 A)** and 3'DDTal2zsmoc1 **(Fig. 4.2.11 B)** were translated successfully, as shown by the single bands that have migrated in accordance with their predicted MW (5'DDTal1zsmoc1 117.16 kDa and 3'DDTal1zsmoc1 117.16 kDa respectively).

To facilitate greater redundancy a 5'Tal1zsmoc1 and 3'Tal1zsmoc1 were both constructed in parallel in the pTAL4 backbone. However, this backbone failed to produce a clean translation product for either the 5' or the 3' constructs. Since single bands were not detected by Western blot these constructs were not carried on for further use **(Fig. 4.2.11 B).**

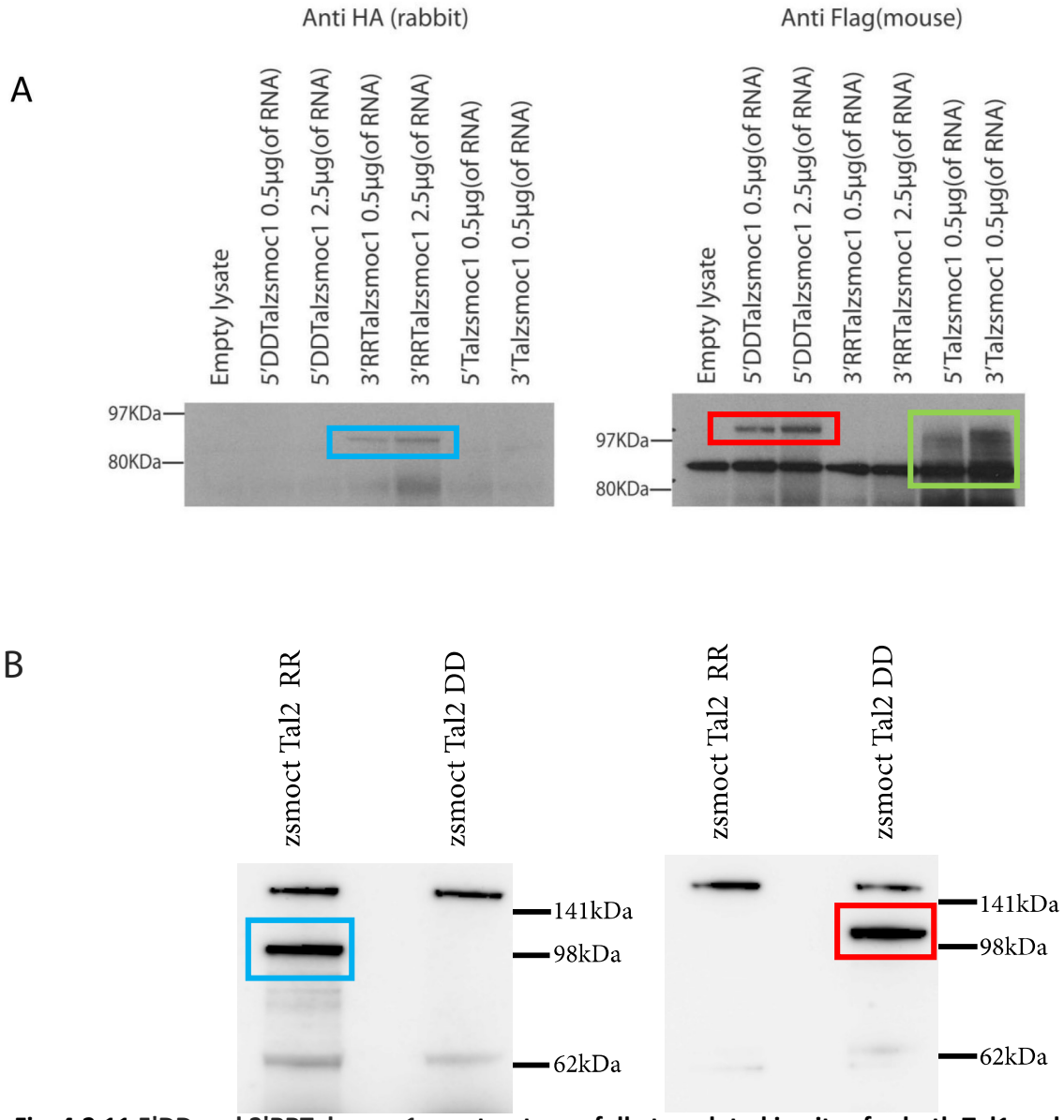



**Fig. 4.2.11 5'DD and 3'RRTalzsmoc1 constructs are fully translated in vitro for both Tal1 and Tal2.**

mRNA was initially transcribed using the mMessage mMachine kits (Ambion), transcripts were tailed with Poly(A) Tailing Kit (Ambion) before finally being translated with the Retic Lysate IVT™ Kit (Ambion) translation then were confirmed by the Western blot. (A) Initially the membrane was probed with anti-FLAG antibody which confirmed that the 5'DDTAl1zsmoc1 had been correctly translated (red box), this Western also shows the in vitro of the same TALEN pair cloned into the TAL4 backbone, this resulted in a smear (green). The membrane was then stripped and probed with anti-HA, unique bands were only present in the synthesis reactions of 3'RRTAl1zsmoc1 plasmid (103kDa). Anti-FLAG antibody which confirmed that the 5'DDTAl1zsmoc1 (red box), the Western also shows the in vitro of the same TALEN pair cloned into the TAL4 backbone, this resulted in a smeared band (blue box). The membrane was then stripped and probed with anti-HA, unique bands were only present in the synthesis reactions of 3'RRTAl1zsmoc1 plasmid (103kDa). **(B)** First the membrane was probed with anti-FLAG antibody which confirmed that the 5DD'TAl2zsmoc1 (red box) membrane was then stripped and probed with anti-HA, bands were only present in the synthesis reactions of 3'RR talent backbones (117.6kDa) and 3'RRTAl1zsmoc1 control.(western blots in B were taken from fig 4.2.17.)

Southern blotting was used to assess the cleavage activity of both TALEN pairs. Initially there was no information on which buffers are optimal for TALEN activity in vitro. So a selection of commercial buffers which were recommended for use for FokI restriction digests were used in conjunction with a mixture of 5'DDTal1zsmoc1 and 3'RRTal1zsmoc1 **(Fig. 4.2.12 A).** Under these conditions DNA cleavage was observed in the samples that were incubated with the TALEN pair and the target DNA. The most pronounced TALEN activity occurred in the reaction using Roche buffer A. To ensure that it was the presence of both the 5' and 3' TALENs that lead to cleavage, a further digest was performed but with additional control digests containing either 5' or 3' targeting proteins but not both **(Fig. 4.2.11 B).** Again cleavage was only observed in the sample treated with both TALENs in the pair. This confirmed that both TALENs were necessary but not sufficient to induce double-stranded breaks in target DNA.

The same protocol was repeated for 5'RR Tal2zsmoc1 and 3'DDTal2zsmoc1 **(Fig. 4.2.11 C).** However, no cleavage products were detected when using this pair.

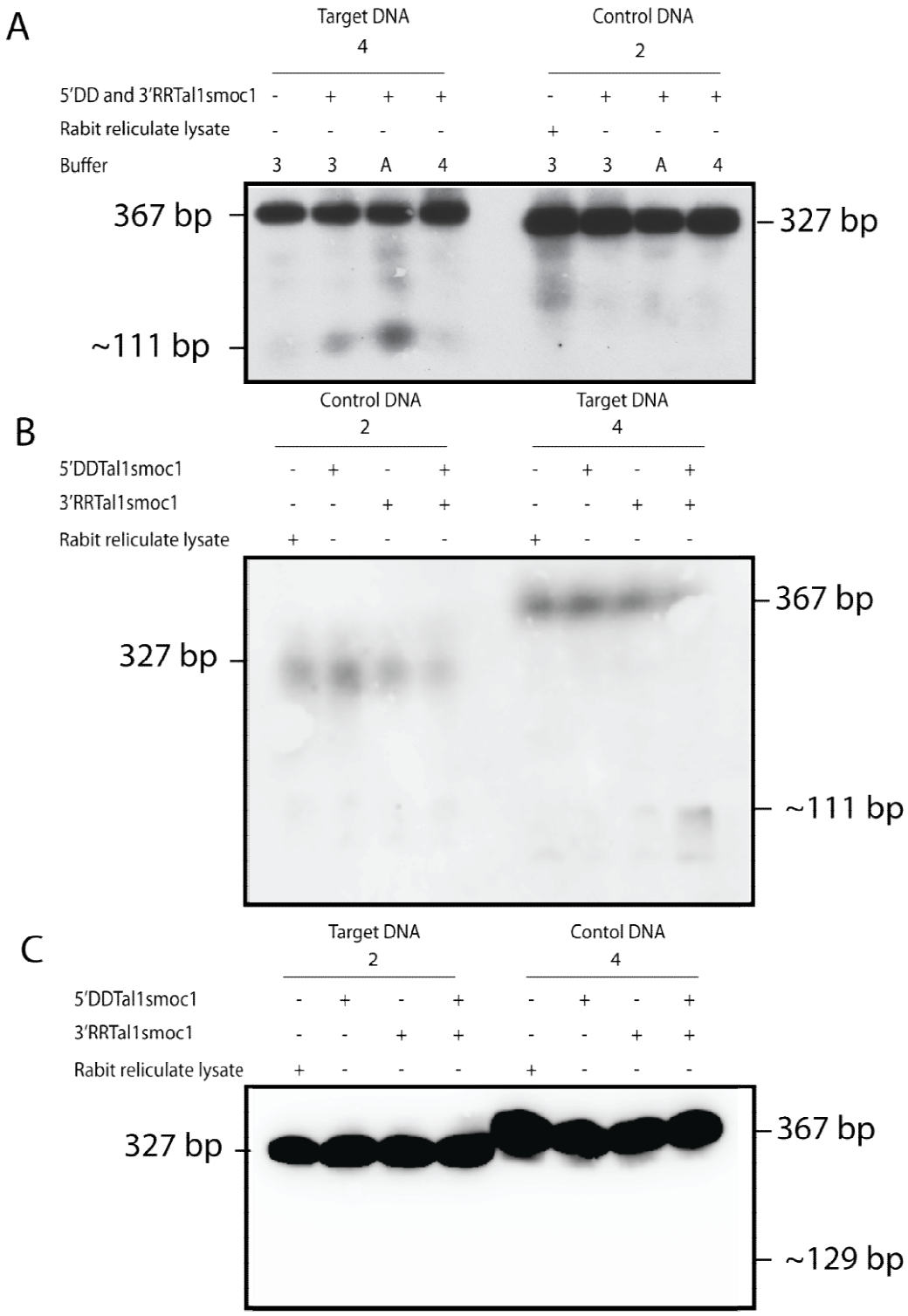



**Fig. 4.2.12 Southern blot analysis of the DNA cleavage activity in vitro.**

**(A)** Shows the effect of different buffer conditions(NEB buffer 3, NEB buffer 4 and Roche buffer A) on the cleavage activity of 5'DD and 3'RRTal1zsmoc1. Cleavage was detected in both NEB buffer 3 and Roche buffer A. As it was highest in buffer A, all further TALEN digestions were carried out in buffer A. **(B)** 5'DD and 3'RRTal1zsmoc1 only cleave DNA when both incubated with the target DNA sequence. **(C)** No cleavage products were detected with DNA incubated with 5'DD and 3'RRTal2zsmoc1.

## 4.2.4 Assessing the endonucleic activity of TALEN pairs 1 and 2 on *zsmoc1* in vivo

Once the TALEN pairs were tested in vitro it was necessary to determine the toxicity of TALEN coding mRNA for zebrafish. Since the aim was to generate the highest cleavage rate possible it was necessary to establish the highest dosage to obtain this, while avoiding unacceptably high mortality. The initial work was conducted using the 5'DD and 3'RRTal1zsmoc1 pair as this pair had been shown to work in vitro **(Fig. 4.2.13 A and B)**. Zebrafish embryos were scored on a three point scale quantifying the number of fish which appeared phenotypically normal, malformed or dead at 24 hpf. All embryos injected showed acceptable levels of mortality and morbidity (<50%). However, there may have been an increase in deformed embryos when injected with 2000 ng/µl. These levels of toxicity are broadly consistent with those reported in the literature for use of TALEN mRNA and superior to those expected from the use of zinc fingers (Cade et al. 2012). As 2000 ng/µl of TALEN mRNA was the highest concentration that could be produced, and the morbidity was acceptable even at this dosage, it was decided that wherever possible the maximum concentration of TALEN mRNA should be injected into the zebrafish embryos .

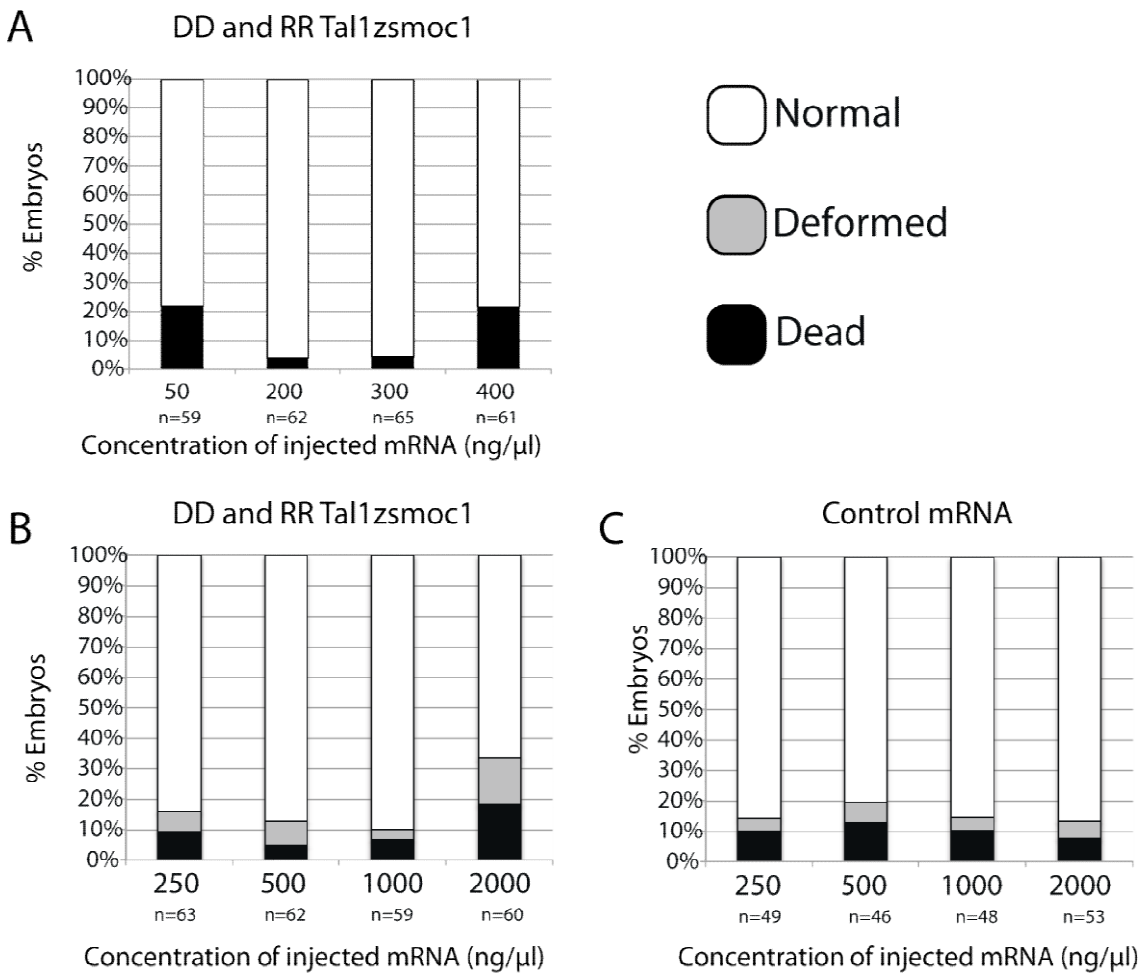



**Fig 4.2.13 The effect of mRNA concentration on the toxicity of TALEN pair injection.**

The percentage of dead and deformed embryos at a given concentration of either DD and RRTal1zsmoc1 TALEN pair **(A)(B)** or a control injection(of antisense RRTal1zsmoc1 TALEN mRNA) **(C)** 24 hpf. TALEN pairs injected at 2000 ng/µl showed an increased toxicity.

n=number of embryos iniected

---

Even the most effective TALEN pairs would be expected to produce a mosaic animal with a heterogeneous population of mutations. Therefore, Sanger sequencing would be ineffective. Clonal analysis has been used to address this (Moore et al. 2012; Cade et al. 2012), whereby PCR products of target DNA are sub cloned into plasmids and a large number of clones are sequenced individually. As multiple clones must be sequenced to estimate cleavage efficiency(which can be as

low as 2%; Cade et al. 2012) this process can be expensive and time consuming when validating large numbers of TALEN pairs. However, next-generation sequencing presents the opportunity to process hundreds of TALEN pairs simultaneously.

It can do this because Next Generation sequencing involves using a semiconductor. Each chip is covered in millions of wells containing a single bead. As long as the DNA sample is sufficiently diluted, only a single DNA fragment will bind to each bead. This is then amplified to produce identical copies. Each bead produces a single independent read, which means an Ion Torrent chip can be used to produce over 80,000 independent reads equivalent to the sequence produced by a single colony in traditional clonal analysis(Merriman et al. 2012).

Next Generation sequencing is most commonly used to process a large number of unique reads. However, to test the TALENs cleavage efficiency we only wanted to capture the specific target site. To this end, two sets of primers were designed to flank both the target sites of both TALEN pairs (1 and 2). One of the disadvantages of current generations of Ion Torrent sequencing is that only relatively short fragments can be sequenced. To account for this primer pairs were designed to be at most 100 bp apart. At least one of the flanking primers was designed to target intronic DNA so the samples could only be amplified from gDNA. gDNA was selected over cDNA because mutated transcripts of mRNA could be lost to nonsense-mediated decay (NMD) or alternative splicing(Lewis et al. 2003; Vaillend et al. 2010).

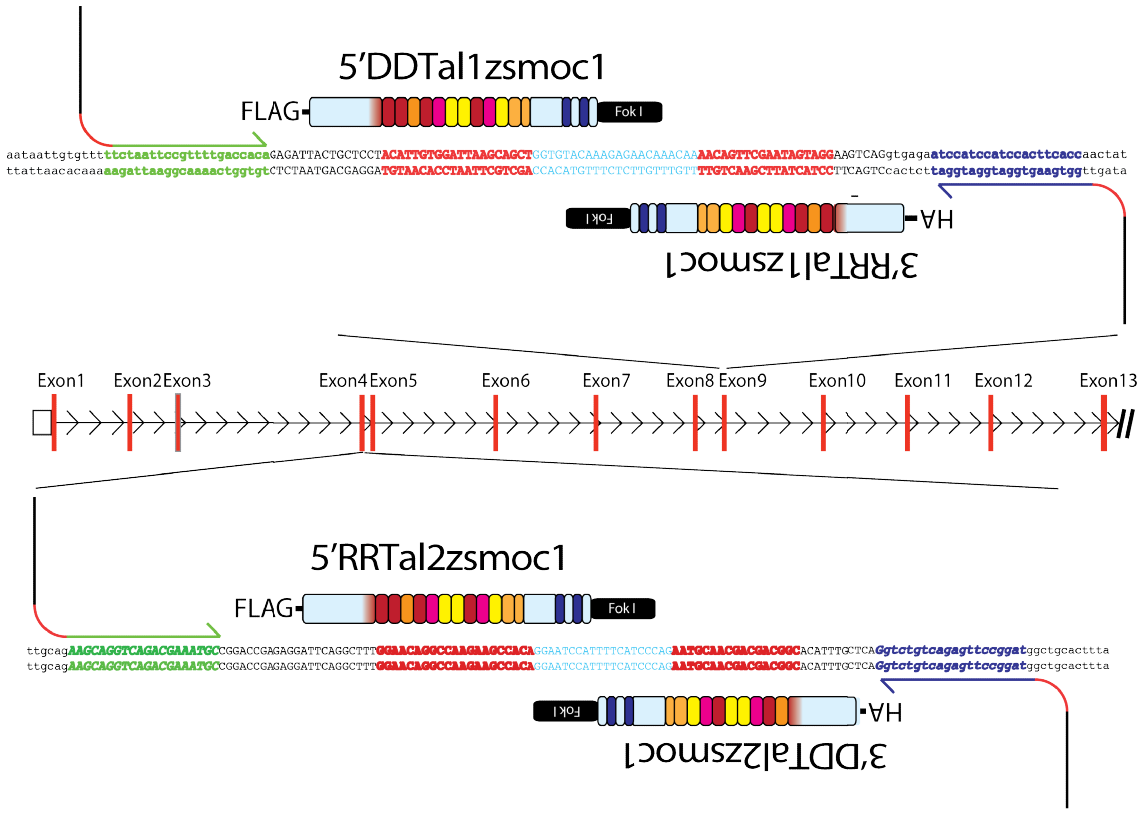



**Fig 4.2.14 Isolating the target site by PCR.**

Primers were designed to flank 100bp of the target exon. Centering on the spacer region (Teal) between the two TALEN binding sites (red). To ensure that only genomic DNA was amplified at least one primer (green = 5' and blue = 3') was located within an intron.

Intronic DNA= lower case  Exonic DNA= uppercase

---

Two types of Ion Torrent primers are commonly used. One is a barcode primer, made up of invariant sequence flanking a 10 bp unique barcode sequence followed by an invariant GAT sequence. The barcode can be recognised by software which then can sort every read with the corresponding barcode. The invariant GAT that follows the barcode helps the software recognize each barcode sequence. The other type of primer is the base primer. This primer contains no barcode, only the target sequence and an invariant sequence. In this case the invariant sequence is used to bind to the bead(Merriman et al. 2012).


**Fig 4.2.15 Ion Torrent Primer sequence designed for *zsmoc1***

The structure of the primers engineered for Ion Torrent sequencing. The target specific base pairs are highlighted in green (forward primers) and blue (reverse primers ) while invariant sequence is highlighted are in black.

To enable the parallel processing of duplicate samples five forward and five reverse primers were designed for each target site. The primers contain a 9 bp unique bar-coding sequence (red). Both primers include invariant sequence used in Ion Torrent sequencing (black). The uses of 10 primers for each target site enables the amplification of eight different PCR products for a single target region with four different barcodes in both the forward and reverse orientation.

As Ion Torrent sequencing produces many orders of magnitude more sequencing data than traditional clonal analysis it was important to establish which were real sequence variants and which sequencing artifacts. To do this we came up with a list of several objective criteria to filter out artifacts. The first step was to quantify the total number of unique deletions detected. Although it is possible for TALEN pair injection to result in other types of mutations, by far the most common mutation arising from the injection of a single TALEN pair are deletions resulting from an error in NHEJ (Mashimo et al. 2013;Kim & Kim 2014). Insertions and missense mutations were considered spurious.

Next we further subdivided between those deletions located within the spacer region of the TALEN pairs targeting site and those located outside the TALEN binding site. Sequencing artifacts should be randomly distributed across the amplified region while legitimate mutations, caused by the combined activity of Fok1 domains, should be confined primarily to the spacer region.

The deletions were then screened by abundance, legitimate mutations were only considered for further analysis if they were present in at least 100 reads.

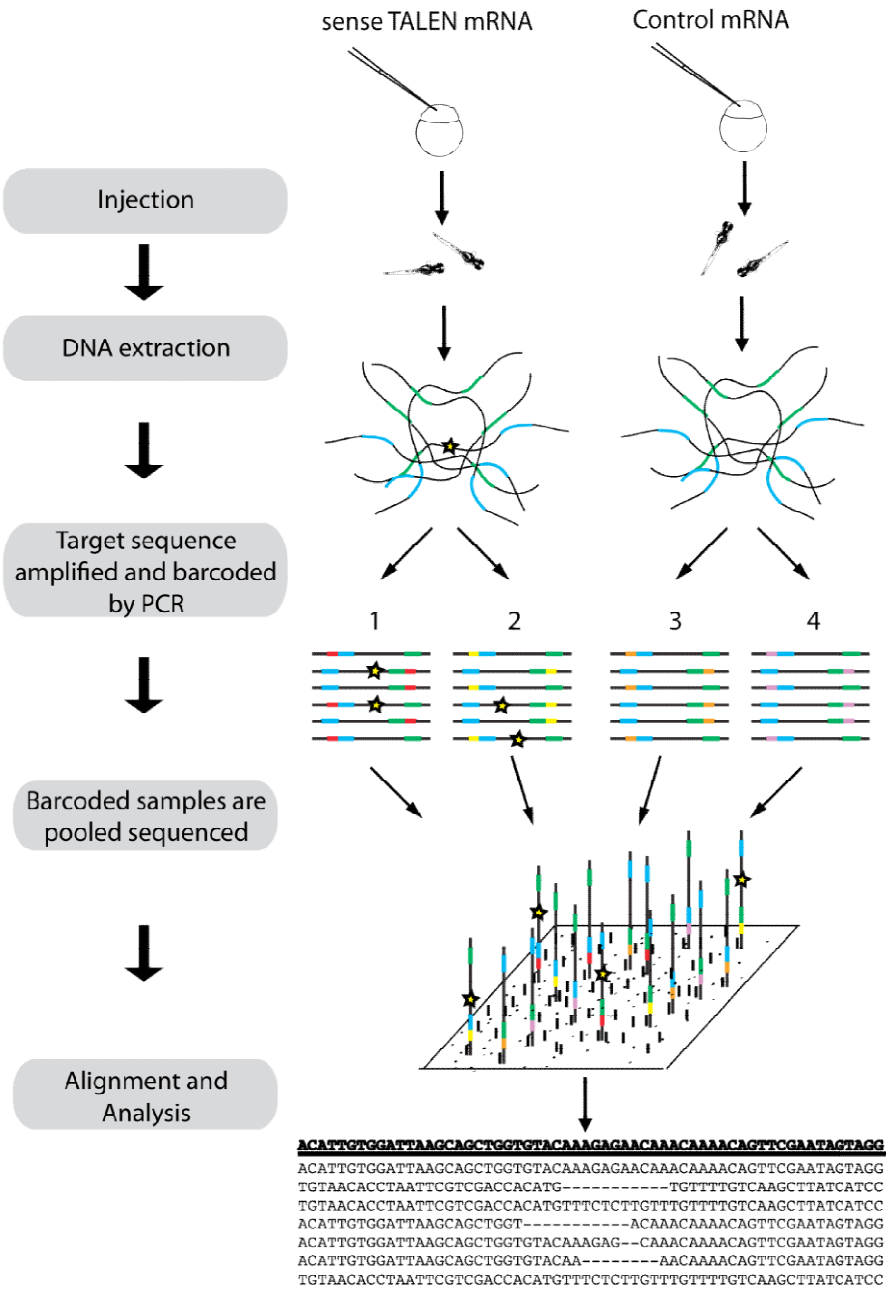



**Fig 4.2.16 A schematic representation of the mutation detection strategy followed using Ion Torrent sequencing.**

Embryos were injected with either a pair of TALEN mRNAs or antisense TALEN mRNA. Embryos were allowed to develop for one day before the genomic DNA was extracted from a small population of embryos. Each sample was then amplified using four sets of primers with two unique barcodes. The PCR product was cleaned and quantified using a bioanalyzer before being diluted to 20 pmol and pooled into one library. Samples were then sequenced in parallel. The sequencing data was then aligned to the target site in NEXTgene and deletions were identified.

None of the tests would distinguish errors introduced during the PCR amplification of the genomic DNA; to filter these PCR artifacts each sample was amplified four times by PCR. Twice in the forwards direction and twice in the reverse. A real deletion should be present in both forward and reverse directions in roughly balanced ratios. Only deletions present in ratios between 1:3 - 3:1 were selected. Not only did this help screen out PCR artifacts (as the forward and reverse reads are amplified independently) but it is also a powerful tool to eliminate sequencing artifacts. The second test is to repeat the experiment using a separate barcode to act as a technical replicate. Real deletions should be present in both populations.

Initially as a pilot study for the Ion Torrent process, eight embryos were injected with 5'DD and 3'RRTal1zsmoc1 for Ion Torrent sequencing **(Fig. 4.2.17 A).** The first sequencing run was successful, but showed no enrichment of deletions in 5'DD and 3'RRTal1zsmoc1 injected embryos when compared to those injected with antisense 3'RRTal1zsmoc1 mRNA. It was noticed that several sequencing artifacts were identified. Ion Torrent sequencing showed that there were inserted thiamine residues in the sequence, although these bases weren't detected in Sanger sequencing of the PCR product. To eliminate these artifacts the screening procedure was modified to include a reverse read, and in all future experiments each sample was amplified with two barcodes.

**A**

| mRNA Injected | 5'DD and 3'RR Tal1zsmoc1 | | Control mRNA |
|---|---|---|---|
| **Concentration of mRNA injected (ng/µl)** | 800 | 2000 | 2000 |
| **Total deletions** | 19 | 24 | 24 |
| **Within target region** | 5 | 6 | 6 |
| **100< reads** | 4 | 4 | 5 |

**B**

| | 5'DD and 3'RR Tal1zsmoc1 | | Control mRNA | |
|---|---|---|---|---|
| **Barcode** | **1** | **2** | **3** | **4** |
| **Total deletions** | 22 | 24 | 19 | 24 |
| **Within target region** | 5 | 6 | 5 | 6 |
| **100< reads** | 4 | 5 | 4 | 4 |
| **Balanced Ratios** | 2 | 2 | 1 | 1 |
| **Shared deletions** | 0 | | 0 | |

**C**

| | 5'DD and 3'RRTal2zsmoc1 | | Control mRNA | |
|---|---|---|---|---|
| **Barcode** | **1** | **2** | **3** | **4** |
| **Total deletions** | 16 | 22 | 23 | 24 |
| **Within target region** | 4 | 3 | 4 | 3 |
| **100< reads** | 3 | 2 | 3 | 2 |
| **Balanced Ratios** | 2 | 2 | 1 | 1 |
| **Shared deletions** | 1 | | 1 | |

**Fig. 4.2.17 No novel mutations were found in *zsmoc1* in embryos treated with *zsmoc1* targeting TALEN pairs**

Ion Torrent sequencing was used to deeply sequence the target region of *zsmoc1*. To rule out sequencing errors sequences were assessed on a number of characteristics. Firstly whether they were deletions. If these deletions were located within the target site. If more than 100 reads(>0.005%) were identical. If the mutations occurred in balanced ratios between fw and rw read. And if these mutations appeared in both barcoded samples. **(A)** Shows the first Ion Torrent sequencing of eight embryos injected with either 800 or 2000 ng/µl of 5'DD and 3'RRTal1zsmoc1 as compared to 8 embryos injected with 2000 ng/µl of antisense control RNA. **(B)** and **(C)** show the results of an Ion Torrent sequencing run of two separate pairs in embryos. **(B)** One injected with 20 embryos 2000 ng/µl 5'DD and 3'RRTal1zsmoc1 and 20 embryos 2000 ng/µl of antisense control RNA and the other **(C)** injected 20 embryos with 2000 ng/µl 5'DD and 3'RRTal2zsmoc1 and 20 embryos 2000 ng/µl of antisense control RNA. Every sample was amplified twice (using two separate barcodes) from gDNA to produce technical replicates. No enrichment of Indel's was detected in any of the sequencing runs.

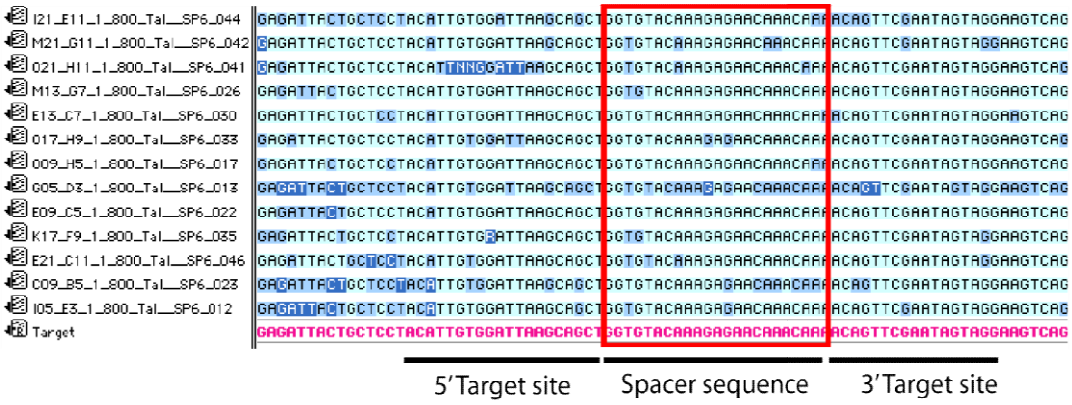
Since the initial experiment was unsuccessful, the injection of 5'DD and 3'RRTal1zsmoc1 mRNA was repeated. This time the number of embryos was increased to 20. Additionally, 20 embryos were injected with 5'RR and 3'DDTal2zsmoc1 and processed simultaneously despite an absence cutting activity in vitro **(Fig. 4.2.14)**. Both batches of embryos were compared to 20 embryos injected with antisense3'RRTal1zsmoc1 mRNA **(Fig. 4.2.17 B,C).**

There was no enrichment in the deletions detected in any of the embryos injected with TALEN pairs, indicating that neither TALEN pair has in vivo endonuclease activity.

Of the 241 mutations identified the screening criteria managed to eliminate over 98% as PCR artifacts.

Although the Ion Torrent sequencing shows that neither TALEN construct has cleavage activity in vivo, it is a new method of assessing endonuclease activity. The possibility remains that the method is ineffective. To assess this traditional clonal analysis was performed **(Fig. 4.2.18 A and B)** on the PCR product amplified in the first deep sequencing experiment **(Fig. 4.2.17 A).** The fragments were TA cloned into the pGMT easy plasmid and DH5 alpha E. coli were transformed. Of the 96 colonies picked for each sample at least 80 were successfully sequenced. The reads were first screened on whether they contained any deletions at all. Then they were screened on whether the mutations were present in both the forward and reverse reads for that colony. No deletions were identified that passed this criteria **(Fig. 4.2.18 B and C).**

A




B

| mRNA Injected | Control mRNA | 5'DD and 3'RRTal1zsmoc1 Pair 1 | |
|---|---|---|---|
| **Concentration of mRNA injected (ng/µl)** | 2000 | 800 | 2000 |
| **Sequenced exon** | Exon 8 | Exon 8 | Exon 8 |
| **Total reads** | 163 | 160 | 162 |
| **Total deletions** | 0 | 0 | 1 |
| **Deletions present in fw and rv reads** | 0 | 0 | 0 |
| **Deletions in target region** | 0 | 0 | 0 |

C

| mRNA Injected | Control mRNA | | 5'DD and 3'RRTalzsmoc1 Pair 1 | Pair 2 |
|---|---|---|---|---|
| **Concentration of mRNA injected (ng/µl)** | 2000 | | 2000 | 2000 |
| **Sequenced exon** | Exon 8 | Exon 3 | Exon 8 | Exon 3 |
| **Total reads** | 162 | 159 | 163 | 160 |
| **Total deletions** | 0 | 2 | 0 | 1 |
| **Deletions present in fw and rv reads** | 0 | 0 | 0 | 0 |
| **Deletions in taget region** | 0 | 0 | 0 | 0 |

**Fig 4.2.18 Clonally analysis of the cleavage activity of *zsmoc1* showed no indels as a result of TALEN Treatment.**

The target exon was amplified using the ion torrent primers detailed in Fig 3.9 however the PCR product was cloned into the pGMT easy vector and DH5 alpha E. coli were transformed. 96 colonies were picked and sequenced using M13 fw and rv primers which flanked inserts. **(A)** A representative screen capture of an alignment of zsmoc1 exon8 the sequence of 13 clones from the gDNA of embryos injected with mRNA (and a concentration of 800 ng/µl) coding for 5'DD and 3'RRTal1zsmoc1. **(B)** the results of the cloning performed using the same PCR product used in Fig A. **(C)** The results of an independent injection of eight embryos with 5'DD and 3'RR performed by Witold Rybski. In this series of experiments all the steps were performed by Witold Rybski, including mRNA synthesis, and cloning.

To test whether there was some unidentified effect caused by my handling of the mRNA or processing of the samples, Witold Rybski repeated the entire procedure for both TALEN pairs from transcription through to clonal analysis. No deletions were present in the forwards or reverse reads obtained **(Fig. 4.2.18 C)** confirming that neither TALEN has in vivo DNA cleavage activity in zebrafish.

To ascertain whether the failure of either TALEN pair to induce any deletions in vivo was caused by the failure mRNA translation SDS PAGE and Western blot was performed on 30 embryos injected with either 5'DD and 3'RRTal1zsmoc1 or 5'RR and 3'DDTal2zsmoc1 against either the HA tag present on the RR vector or the FLAG tag present on the DD vector **(Fig. 4.2.19).** The Western blot confirmed that both TALEN constructs had been translated correctly as protein bands were detected at the predicted molecular weights. There was an unexplained band present in both the in vitro translation and the 5'RR and 3'DDTal2zsmoc1 injected embryos which had an estimated molecular weight greater than 141KDa.

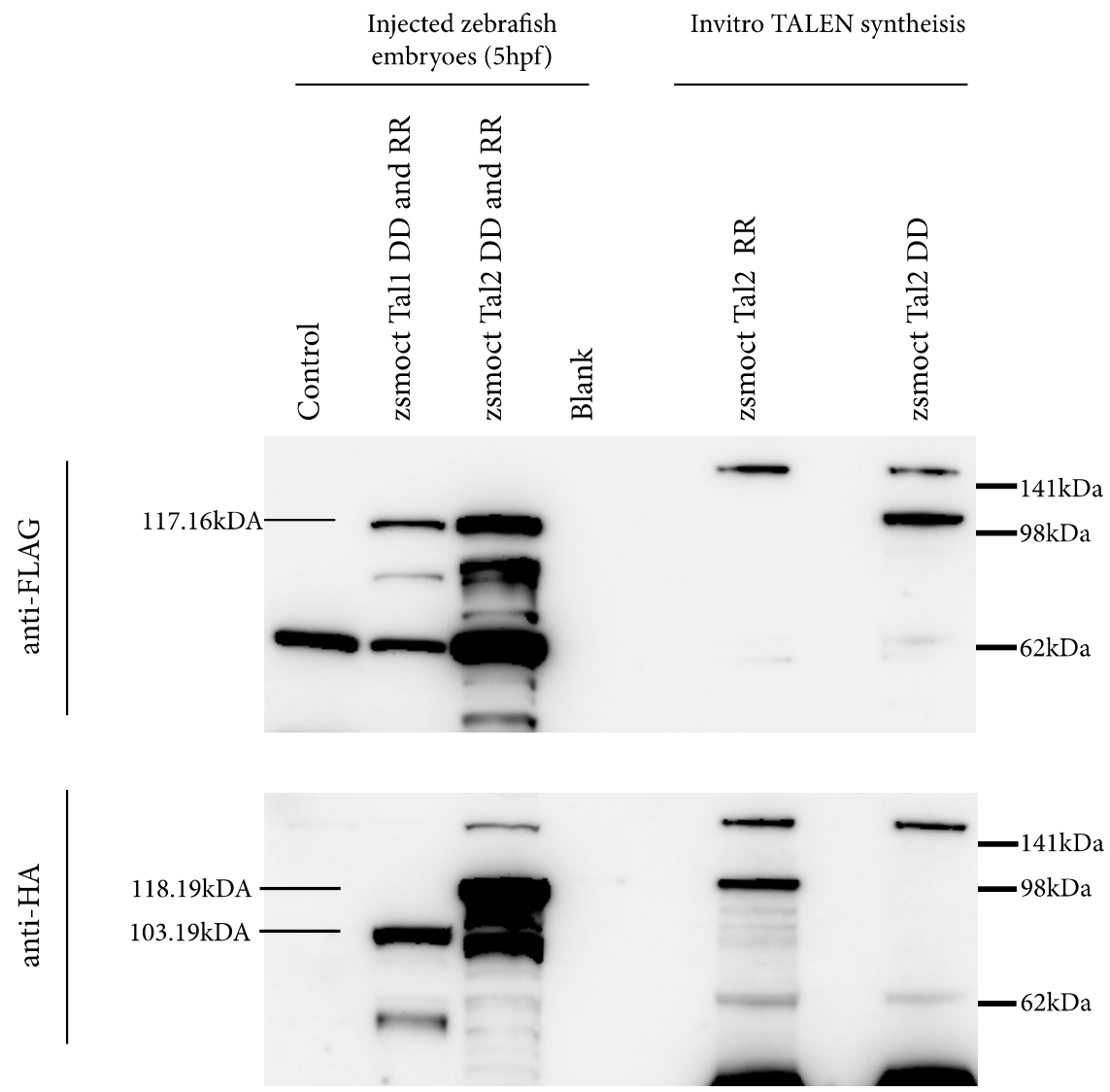



**Fig. 4.2.19 Both TALEN constructs are expressed in vivo.**

The Western blot was probed with anti-HA, which was then stripped and reprobed with anti-FLAG antibodies. The zebrafish was injected at the one cell stage of development and harvested at 6 hpf. As a positive control mRNA 5'RRTal2zsmoc1 and 3'DDTal2zsmoc1 were translated in vitro in separate reactions. The samples were resolved on 10% polyacrylamide gels and the blots were imaged on a ImageQuant LAS.

# 4.3 Discussion

## 4.3.1 Neither TALEN pair produced cleavage in vitro.

The TALEN pairs synthesised showed no activity in vivo despite 5'DD and 3'RRTal1zsmoc1 showing promising activity in vitro. While TALEN pairs usually show some cleavage activity it has been estimated that one in three TALENs fail to show any cleavage in their target sites (Cade et al. 2012). It is puzzling that the TALEN pair 1 showed cleavage in vitro but not in vivo, but there are several possible explanations for this. The first is that the TALENs were not correctly assembled. However, sequencing of both TALEN pairs confirmed that they were assembled correctly and free of mutations **(Fig. 4.2.4; Fig. 4.2.5; Fig. 4.2.6; Fig. 4.2.7; Fig. 4.2.8; Fig. 4.2.9).** A ligation based strategy was used, so spontaneous mutations are unlikely. 5'DD and 3'RRTal1zsmoc1 showed in vitro cutting activity which can only occur if they were correctly assembled **(Fig. 4.2.12 A, B).** Whilst it cannot be ruled out that there was an undetected defect in either 5'RR or 3'DDTal2zsmoc1, it seems unlikely given cleavage activity, the sequencing data and size of the fragments produced by restriction digests.

The second possibility that can be ruled out is that the TALEN mRNA is simply not expressing the TALEN protein. Both the in vitro and in vivo translation

of every TALEN construct was confirmed by western blot **(Fig. 4.2.11; Fig. 4.2.19).** The molecular weights of the detected protein bands also matched those predicted by sequence composition further supporting that the constructs were assembled correctly.

The final possibility that has been ruled out is that the gDNA sequence of the target site in the zebrafish injected with TALEN mRNA, did not match the sequence that the TALEN pairs were designed to target. A single mismatched base pair within the target site has been shown to lead to a fivefold reduction in cleavage activity for mammalian in cell lines (Zhang et al. 2011). In live zebrafish a 2 bp polymorphism in the *epas1b.2* target site led to a decrease from cleavage efficiencies of 76% and 56% to 2% (Cade et al. 2012). This is conclusively ruled out in this case because when assessing cleavage in vivo target sites both TALEN pairs were either sequenced by Ion Torrent or Sanger sequencing to identify deletions, and the sequences matched those used to design the TALEN pairs **(Fig. 4.2.17).**

It is therefore likely that the lack of mutations is caused by some unknown context dependent effect. While the TALE code is often presented as a direct RVD to base code it is worth noting that in nature there is no single TALE target site which directly corresponds to this code (Boch et al. 2009; Cong et al. 2012). Although TALE technology is a new technology it has already been shown that RVD order and composition do affect TALEN activity (Christian et al. 2012 ; Meckler et al. 2013).

DNA methylation at the target site is also known to significantly inhibit TALEN cleavage (Bultmann et al. 2012). This is due to HD subunits having limited

binding specificity to methylated cytosine, although proprietary subunits do exist that have specificity for methylated cytosine base pairs. The methylation status of the target site is currently unknown. However, it is unlikely that methylation is the causative factor because the target site is located in the centre of the *zsmoc1* gene, and faraway from the known promoter regions of any known genes (Brenet et al. 2011).

There are also several factors which have not been tested but are plausible candidates for modifying TALEN activity. These include transcriptional state, binding partner and DNA confirmation.

Another possibility is that the TALEN pair is functional in vivo but that it has a low cutting efficacy. In the in vitro system when the TALEN pair cuts the target DNA there is no possibility of repair and homologous recombination. However, when DNA is cleaved in vivo it is repaired. In fact, the ability of TALENs to induce DNA repair is vital to their efficacy (Anon 2011; Bogdanove & Voytas 2011). Is been shown that lesions in the DNA caused by endonucleases lead to induction of none homologous end joining (NHEJ), which can lead to errors resulting in micro deletions (Gong et al. 2005). However, not all NHEJ is equally error prone. In fact the classical NHEJ events are remarkably good at resolving the kind of double-stranded DNA breaks resulting from endonuclease activity. These breaks all lead to compatible over hangs which require no further processing before they can be resolved. It has been demonstrated that most of the cuts induced by endonucleases are resolved without any errors (Mladenov & Iliakis 2011). If the targeted TALEN pair cut efficiently enough, it can overwhelm the cNHEJ repair pathway by repeatedly cleaving the same site resulting in the cycle of

repair and cleavage which can only be resolved by mutation. If this is true, then it should be possible to induce mutations by inhibiting cNHEJ and forcing the organism to use a more error prone pathway. This has now been achieved in cell lines (Certo et al. 2012) and in rats (Hwang et al. 2013) by co-expressing a TALEN pair with *Exonuclease 1*(*Exo1*). Exo1 can digest the overhangs left when the TALENs cleave the DNA, which means they are no longer homologous, forcing the cell to use the alternative NHEJ pathways resulting in larger and more frequent deletions (Certo et al. 2012). It would be interesting to see if *exo1* expression leads to an up-regulation in TALEN efficacy in my constructs. It would both suggest a possible reason for the discrepancy between the in vitro and in vivo assays and be a novel finding; as yet there have been no reports of exo1 co-expression in zebrafish to increase TALEN efficiency.

### 4.3.2 Evaluating the novel assays developed to assess TALEN activity

While the in vitro cleavage assay did work in principle and cleavage was detected for the co-injections of 5'DD and 3'RRTal1zsmoc1. This did not translate into in vivo activity. It is questionable whether it is worth pursuing this strategy in the future as yeast reporter systems (Sun, Ning 2013) have now been developed which can rapidly assess TALEN cleavage activity in vivo. The results have clearly shown that TALEN pair's high cleavage efficiencies in yeast are highly likely to have at least some cleavage activity within the desired organism (mammalian cells,

zebrafish, *Drosophila* etc). It is likely that in vivo screens, even between distantly related organisms, better reflect the conditions in the desired target species.

The Ion Torrent approach developed holds more promise. Modern manufacturing processes can produce hundreds of TALEN pairs a day which enables their use in large-scale experiments. For example TALEN pairs have been produced targeting every known gene and several hundred micro RNAs in the human genome (Kim et al. 2013). These pairs still need to be validated to confirm that they induce deletions in vivo. Clonal analysis is obviously less than ideal. Even if CRISPR technology completely supplants TALEN technology the need to validate large cohorts for deletions remains the same. The Ion Torrent sequencing approach has the advantage of being able to process multiple TALEN pairs simultaneously. Using the 96 barcodes validated by life sciences it would be easy to process 24 TALEN pairs simultaneously, and it is easily possible to generate custom barcodes as needed (10bp mean there are theoretically $4^{10}$ or 1048576 combinations, though not all sequences are usable). By using a more advanced Ion Torrent sequencer, such as the Ion proton there would be no need to even sacrifice depth.

Neither TALEN pair actually worked in vivo so it has not been proved that Ion Torrent sequencing can detect TALE induced cleavages. However, the work undertaken is a good first step. The criteria for filtering out PCR and sequencing artifacts worked well **(Fig. 4.2.17).** The clonal analysis confirmed the Ion Torrent results **(Fig. 4.2.18).** Given more time it would be worth repeating the experiments with TALENs of known efficacy to establish if deletions are detectable, and if so optimisation would be highly desirable.

## 4.3.3 Alternatives to a TALEN based strategy

Even though the TALENs based strategy has been unsuccessful it is unlikely that switching to a ZFN based strategy would be successful. Studies comparing ZFN activity to TALENs have consistently shown them to be inferior to TALENs. They have fewer target sites, (Sander et al. 2011) show more off target effects (Mussolino et al. 2014; Ding et al. 2013) and lower cleavage efficiencies (Cade et al. 2012; Chen et al. 2013). Until recently ZFN and TALEN technology have been the only options available for forward genetics in zebrafish, but towards the end of this project Clustered Regularly Interspersed Short Palindromic Repeat (CRISPR) technology has been making waves in the field of programmable nucleases (Hwang et al. 2013). CRISPRs function in a completely different way to both TALENs and ZFNs. The proteins form part of the adaptive immune system of bacteria and archaea (Jinek et al. 2012; Wiedenheft et al. 2012). When the microbe is infected by plasmids or phages the DNA can be incorporated into specific spacer regions of these organisms genome. The sequence is then transcribed and processed, forming short (~20bp) specialised CRISPR RNAs (crRNAs). It is these crRNAs that target the DNA cleavage. The nuclease is coded for by Cas proteins (typically Cas9 for genomic engineering) which bind to the crRNAs. These allow the Cas proteins to recognise the target site and induce double strand DNA breaks **(Fig. 4.3.1;** Blackburn et al. 2013).

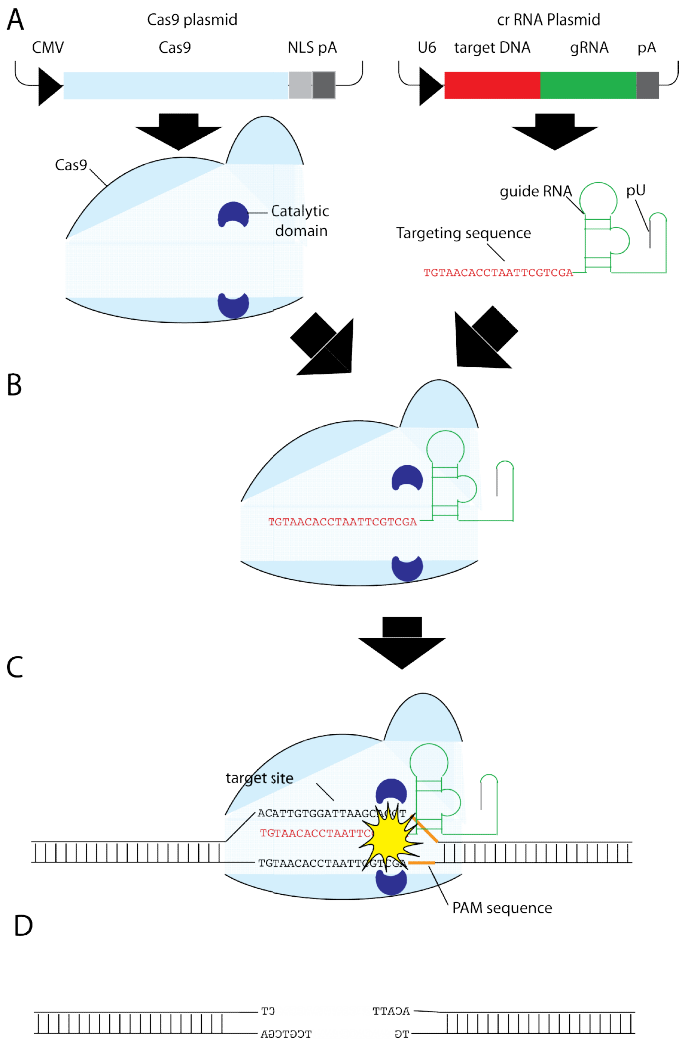



**Fig 4.3.1 A schematic representation of a type II CRISPR system.**

**(A)** Initially two plasmids must be used to generate both the Cas9 protein (which provides the cleavage activity) and the crRNA (which provides the targeting specificity). **(B)**The protein and crRNA are mixed in vivo. The Cas9 protein forms a complex with the crRNA by binding to the guide RNA of the crRNA. **(C)** The complex can then recognise the target site by the complementary base pairing of the targeting sequence as long as the targeting sequence has a 3' PAM sequence. The nucleases domains cleave the DAN approximately 3 base pairs upstream of the PAM sequence. **(D)** Which creates double strand breaks that can induce mutations through NHEJ.

Figure drawn from information found in Blackburn et al. 2013 and Mali et al. 2013.

---

Recently it became possible to engineer artificial crRNAs, which can be used to target the desired DNA sequence. The principal advantage of the system is its simplicity. *Cas9* mRNA can be directly transcribed from a standard plasmid while the crRNA mRNA can be synthesised by ordering two overlapping PCR primers which can then be cloned into a guide RNA expression vector. Knockouts have already been generated in many model organisms including zebrafish (Hwang et al. 2013; Chang et al. 2013; Hruscha et al. 2013; Ota et al. 2014; Jao et al. 2013). While comprehensive studies of CRISPR induced mutations have yet to be

published (publication leads to bias for successful mutations) CRISPR cleavage activity seems to be much higher than that of most TALEN or ZFN pairs with cleavage efficiencies generally in the region of 75–90% (Jao et al. 2013). This activity is often high enough to produce biallelic mutations with high enough frequency to induce $F_0$ phenotypes. It has even been demonstrated that CRISPRs can produce multiple targeted deletions in both mice and zebrafish in one generation (Ota et al. 2014; Wang et al. 2013). Jao et al. managed to knock down tyr, golden, mitfa, and ddx19 along with a GFP reporter and showed that a combined corresponding phenotype of four of the mutants was observed in individual $F_0$ fish (Jao et al. 2013). This opens the way to much faster experiments on combinational knockouts. In fact, the CRISPR system is well-suited to multiple knockouts; its modular nature allows injection of one set of CAS9 mRNA along with multiple short targeting RNAs.

There are downsides to CRISPR technology. While the PAM is 20bp long CRISPRs have been shown to tolerate up to 5 bp miss matches in vivo (Fu et al. 2013), so the actual targeting domain is short (14bp). This means the chance of off target effects is comparatively high (Wiedenheft et al. 2012). This is compounded by the fact that because the nuclease is within the Cas protein it's impossible to separate the nuclease domain and use the CRISPRs in pairs as you can with both TALEN and ZFN technology. Additionally, there is a requirement that the target site be preceded by a protospacer adjacent motif (PAM) sequence which means genes have fewer acceptable target sites than a TALEN pair (Blackburn et al. 2013). Mutations induced by CRISPR tend to be much less diverse than those induced by either TALEN or ZFNs because they bind to the target site they target.

Which means that once a mutation has been induced it cannot be re-mutated. This is a double-edged sword. If the aim is to produce consistent $F_o$ phenotypes it is desirable. However, if you want to obtain an extensive allelic series it is detrimental. These problems may be being close to being resolved as recently a novel strategy using paired CRISPR nucleases has been reported (Ran et al. 2013). Ran et al. used two mutated Cas backbones (Cas9n and Cas9H840A) that induced a single strand DNA nick though on different strands. This wouldn't normally lead to deletions but by using PAMs that target sites close enough together the induced single stranded nicks that can form double stranded breaks. They achieved cleavage efficiencies of greater than 80% in mice, with a 50-1500 fold reduction in off target cleavage (Ran et al. 2013).

Obviously the outlook for CRISPR technology is incredibly positive and had the technology been available earlier in the project it would have been pursued.

### 4.3.4 Concluding remarks

The majority of the aims of this chapter were successfully completed. The TALEN pair were successfully designed, constructed and validated. A novel way of assessing TALEN cleavage efficiency in vitro was developed and a promising new strategy for assessing multiple TALEN pairs was developed.

It is unfortunate that neither TALEN pair produced heritable deletions in the zebrafish. Every effort was taken to verify that they were correctly constructed, but it will probably never be established why the TALEN pairs synthesised do not work in vivo. A *zsmoc1* null zebrafish is still highly desirable to confirm the results of the morpholino experiments. It is possible that CRISPR technology could be used to generate a germline loss of function mutation in the near future.

# Chapter 5

# Discussion

# 5.1 Future work

## 5.1.1 Generating a *zsmoc1*$^{-/-}$ zebrafish lines

Despite the fact that every effort was made to obtain a control for off target effects and morpholino toxicology the availability of efficient targeted endonucleases has further demonstrated that most morphant phenotypes are often unreproducible. A good example of this is the recent work of van Impel et al. on the development of the lymphatic system in zebrafish. They generated knockout zebrafish lines for four genes (*sox18*, *nr2f1a* and *prox1a/b)* all of which had published lymphatic morphant phenotypes. Yet every single line failed to phenocopy the zebrafish morphants, and none of them had a lymphatic phenotype (van Impel et al. 2014). Even the validity of mRNA rescue has been called into question. Whilst it clearly demonstrates that at least part of the phenotype is attributable to the absence of the rescued gene it does not rule out off target effects (Schulte-Merker & Stainier 2014), which is particularly troubling given the partial nature of the rescue of *zsmoc1* morphants with *hSMOC1* mRNA **(Chapter 3 3.2.7)**.

With this in mind it becomes very important to confirm these results using reverse genetics. Targeted endonucleases too can bind to off target sites and produce off target effects, but even at this early stage they appear to be

appreciably less toxic than morpholinos. Even in the cases when they do generate off target deletions within functional genomic regions it is possible to segregate away off target mutations by out crossing the zebrafish line. Usually this can be achieved within two or three generations (Schulte-Merker & Stainier 2014).

The efforts made to address this issue and establish *zsmoc1* knockout lines are detailed in **chapter 4.** Unfortunately these efforts were unsuccessful. Whilst many groups have reported successful use of TALEN pairs (Moore et al. 2012; Cade et al. 2012; Sander et al. 2011) it is clear that TALEN targeting is not as simple as it was first thought to be **(Chapter 4 4.3.1)**. Whilst it may be possible to generate TALEN pairs successfully to target the *zsmoc1* locus, with the availability of CRISPRs it is clear that technology has already moved on. CRISPR technology shares almost all of the advantages of TALEN technology, and it appears to be superior in a number of critical ways.

CRISPRs have been reported to have lower failure rates and higher cleavage efficiencies (Jao et al. 2013; Ota et al. 2014; Wang et al. 2013). The extent of off target effects remain to be fully analysed. The two platforms appear to have comparably low toxicities. CRISPRs often have high enough cleavage efficiencies to generate phenotypes in the $F_0$ generation. When this is the case they can be used in the same way as MO, even with the use of multiple CRISPRs (Jao et al. 2013). Coupled with the fact that CRISPRs are considerably easier to synthesise (it takes just one week, requires just two plasmids and it is possible to sequence the final construct and conclusively eliminate any mutations) (Blackburn et al. 2013 ; Mali et al. 2013). It becomes obvious that CRISPR should be the preferred platform for the generation of knockout lines.

Of course a knock out line is not only useful for verifying the MO phenotype it would also be a valuable model system to study the action of *zsmoc1*, particularly any effects that occur later in zebrafish development. This is because both MOs and mRNA become less active over the course of development and are inactive by 5 dpf (Nasevicius & Ekker 2000; Skromne & Prince 2008).

### 5.1.2 Investigate whether zsmoc-1 is acting on any other signalling pathways

Whilst the link between smoc1 and BMP signalling is now well established there remains the question of whether zsmoc-1 interacts with any upstream zebrafish pathways. There are several plausible candidates, although MAPK signalling is the only candidate with evidence supporting a link with bmp signalling. Thomas et al. demonstrated that while *xsmoc1* mRNA could rescue constitutively active *xbmpr1* in animal caps they could not rescue the phenotype of cell lines with linker mutated (LM)-p-smad1 (Thomas et al. 2009). As LM- smad1 is auto-phosphorylated at a residue under the control of MAPK signalling they interpreted the result is demonstrating smoc1 acted via the MAPK pathway (Thomas et al. 2009), and went on to show that *xsmoc1* morphants displayed significantly increased levels of p-erk. Consistent with this model, excess Smoc-1 has been shown to reduce Smad signalling and increased p-Erk levels in rat tissue culture (Dreieicher et al. 2009).

Map kinase signalling is involved with both the establishment of the dorsal ventral axis in zebrafish. *fgf8* forms a dorsal gradient, and its overexpression suppresses bmp signalling leading to a dorsalised phenotype (Fürthauer et al. 2004). Conversely inhibition of Mapk cascade by overexpressing the Mapk pathway phosphatase *mlp3* leads to the expansion of the expression domains of both bmp7 and bmp2b expression ventralising embryos (Tsang et al. 2004). Fgf signalling also has an established role in vertebrate eye development, including the dorsal ventral patterning of the zebrafish retina (Adler & Canto-Soler 2007).

Alternatively the results in both Thomas et al. and Dreieicher et al. can also be explained if smoc-1 acts via the non-canonical bmp pathway (Hu et al. 2004; Hartung et al. 2006), which is smad-independent and involves signalling via the MAPK pathway.

While there is no evidence for effects on either wnt and shh signalling both are key players in vertebrate eye development and plausible candidates for smoc-1 action (Adler & Canto-Soler 2007).

The live imaging techniques developed in this thesis could easily be used to carry out pilot work on establishing which, if any of the signalling pathways are linked to z=smoc-1. Zebrafish transgenic reporter lines are available for both mapk (dusp6; Molina et al. 2007) wnt (Tcf/Lef; Shimizu et al. 2012) and shh signalling (TOP; Dorsky et al. 2002). All of which express gfp in the optic vesicle and/or retina by 24 hpf (Molina et al. 2007; Shimizu et al. 2012; Dorsky et al. 2002).

## 5.1.3 Establish the protien localisation of zsmoc1

While RNA *in situ* hybridizations of zsmoc1 have shown that smoc1 transcript is present in the zebrafish eye it has yet to be established whether these transcripts were actually translated. Several attempts were made to validate anti-smoc1 antibodies unfortunately all showed significant non specific banding/staining zebrafish(data not shown). Had there been sufficient time a GFP tagged smoc-1 could have been engineered. Not only would a smoc1 reporter line allow the dynamic visulisation of the localisation of smoc1 through development, it would also provide further evidence that morpholino treatment is actually reducing the levels of translated protien.

## 5.1.4 Investigate the link between zsmoc-1 action, bmp signalling and the glypicans

While my results support a model of the relationship between zsmoc-1 and bmp signalling that mirrors that of Pentagone and Dpp. more work needs to be done to confirm this theory. The best way to establish this would be to develop a model system to visualize the gradients in real-time: a system akin to the imaginal disk of the *Drosophila* wing which has been used to great effect to probe the mechanisms of control of Dpp signaling (Vuilleumier et al. 2010; Hamaratoglu et al. 2011). With its ex-vivo development and transparent embryos zebrafish are

ideally suited to being the vertebrate systems of choice for the real-time imaging of signalling gradients.

Our work shows two potential p-smad gradients which could be further developed into the system. The first potential model is the ventral-dorsal p-smad gradient which I have demonstrated is regulated by smoc-1 **(3.2.4 and 3.2.8)**. This developmental stage presents a valuable opportunity to test the prediction that zsmoc-1 action is dependent on the presence of at least one Glypican. This is because the *knypek* is already known to be expressed at the same time and involved in controlling dorsal migration in convergence(Topczewski et al. 2001). Coupled with the fact that both its mutant and morphant phenotypes are very similar to those of the *zsmoc1* morphants is certainly suggestive of a link. However, there are obstacles to using this system. My main objection is that current generations of fluorescent reporter genes do not generate detectable levels of GFP this early in signalling development (Molina et al. 2007; Collery & Link 2011).

Perhaps a better system would be to further investigate the dorsal ventral psmad gradient found in the zebrafish retina. Whilst it's unclear which (if any ) of the six vertebrate glypicans are expressed in the zebrafish retina during dorsal ventral retina patterning, reporters for all the main signalling pathways show detectable levels of expression in the OV (Molina et al. 2007; Shimizu et al. 2012; Collery & Link 2011; Dorsky et al. 2002). What's more, given the severe ocular abnormalities of OAS the precise function of zsmoc-1 in eye development is of particular interest. The first step to establishing this system would be to confirm that the increase in p-smad signalling leads to the predicted expansion of the

dorsal retina. The easiest way to establish this is with the use of situ hybridisation to measure the extent of the expression domains of known retinal fate marker genes (e.g. *bambi*, *tbx5* and *vax2*; French et al. 2009).

Both gene knockouts and MO knockdowns would result in the universal down-regulation of the target gene, which in the case of *zsmoc1* leads to severe embryonic defects which severely affect the eye field. To combat this we used lower doses of zsmoc1tMO. A more elegant solution would be tissue-specific knockouts. CRISPR technology now makes it far easier to develop transgenic zebrafish lines. It is not only easier to create knockouts of choice genes but also to rescue those knockouts with targeted insertion of an identical gene flanked by LoxP sites (Hwang et al. 2013; Wilkinson & Wiedenheft 2014). Any gene flanked by LoxP can then be knocked down in the OV by crossing the LoxP line with a tissue specific CRE line (e.g. In this case *tg(rx3:cre))*. Tissue specific up-regulation can be attained by using transgenic lines that ectopically express the gene of interest within the OV (e.g. T*g(rx3:smad7)*; French et al. 2009).

There are a number of unresolved issues in the general mechanics of signalling gradients formation and maintenance and smoc-1's mode of action which this system could address, but to my mind the most pressing is the role of endocytosis/transcytosis in great formation. Currently this work is almost entirely based on studies conducted in *Drosophila* (Akiyama et al. 2008; Belenkaya et al. 2004; Schwank et al. 2011). However, as the existence of multiple models of gradient formation (e.g. RET and RED) shows the situation is less than clear even in *Drosophila,* and almost no work has been done on real-time quantification of signalling molecules in vertebrate embryos.

# 5.2. Contribution to the field

By definition every Ph.D. thesis must contribute novel knowledge to its field of study. My Ph.D. is no exception. To finish I will summarise the main contributions I have made in ascertaining smoc-1's role in embryonic development while conducting this thesis.

## 5.2.1 Corrected the zebrafish locus of the zsmoc1

The first significant result detailed in **chapter 2** was the correction of the miss annotated *zsmoc1* sequence. I not only proved that the two *zsmoc1* entries recorded in UCSC genome browser are actually part of a single gene. I also showed that the *zsmoc1* was missing several exons and both the 5' and 3' ends of the gene. I went on to isolate the signal peptide sequence through five prime RACE. Now that the actual sequence of *zsmoc1* has been established it should be considerably easier for other researchers to conduct studies on zsmoc-1's function using zebrafish as a model organism.

## 5.2.2 Established novel assays of cleavage efficiencies of targeted in endonucleases

The second main contribution was in chapter 4 with the development of two new assays to assess cleavage efficiency of TALEN pairs. With the advent of TALEN and CRISPR technology targeted endonucleases are going to become increasingly central to modern biology: one which allows the assessment of in vitro cleavage activity using southern blotting and the other using Ion Torrent sequencing to quantify the function in vivo. Of the two the in vivo test is potentially more useful. Enzymes with high in vitro cleavage activity do not seem to reflect in vivo activity, and since the technique was developed yeast hybrid cleavage assays have been shown to make more accurate predictions of in vivo cleavage activity.

## 5.2.3 Provided the first functional work supporting a model where smoc-1 acts as a regulator of gradient shape

The third and my most central contribution to the field of developmental biology is the functional work on zsmoc-1. Not only is this the first in-depth investigation into the functional role of smoc-1 in zebrafish but I'm the first to implicate smoc-1 in the regulation of the convergence movement of the gastrulation zebrafish embryo through the regulation of the lateral but not sagittal elements of the dorsoventral p-smad gradient **(3.2.2; 3.2.19)**. While my results are

inconsistent with those obtained by Abouzeid et al. I have observed an eye phenotype in *zsmoc1* morphants **(3.2.15)**. Though the zebrafish did not display microphthalmia or anophthalmia they did have significantly regulated BRE expression in the dorsal retina. This would likely result in defects in retinal patterning (Asai-Coakwell et al. 2007; Asai-Coakwell et al. 2009; French et al. 2009), though these remain to be observed. Given the significant eye defects associated with OAS a zebrafish eye phenotype increases its usefulness as a model for the disease. While it is nowhere near as severe as the anophthalmia observed in OAS it is worth noting that even the mouse knockouts of *SMOC1* show comparatively mild phenotypes (Rainger et al. 2011; Okada et al. 2011).

My results support the function of smoc-1 as a regulator of bmp signalling and provide the first support for the function of vertebrate smoc-1 being broadly conserved to *Drosophila.* Building on the work of Vuilleumier et al. and Hamaratoglu et al. on *Pentagone* I have proposed a model for the action of the smoc-1 in vertebrates that is both consistent with the results I obtained and those published in *Xenopus* (Vuilleumier et al. 2010; Hamaratoglu et al. 2011: Thomas et al. 2009*;* David Sexton 2010).

# Bibliography

Abouzeid, H. et al., 2011. Mutations in the SPARC-related modular calcium-binding protein 1 gene, SMOC1, cause waardenburg anophthalmia syndrome. *American Journal of Human Genetics*, 88(1), pp.92–98.

Adler, R. & Canto-Soler, M.V., 2007. Molecular mechanisms of optic vesicle development: complexities, ambiguities and controversies. *Developmental Biology*, 305(1), pp.1–13.

Akiyama, T. et al., 2008. Dally regulates Dpp morphogen gradient formation by stabilizing Dpp on the cell surface. *Developmental Biology*, 313(1), pp.408–419.

Anderson, S., 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Research*, 9(13), pp.3015–3027.

Anon, 2011. Move over ZFNs. *Nature Biotechnology*, 29(8), pp.681–684.

Arnold, S.J. et al., 2006. Dose-dependent Smad1, Smad5 and Smad8 signaling in the early mouse embryo. *Developmental Biology*, 296(1), pp.104–118.

Arora, K. et al., 1995. The Drosophila schnurri gene acts in the Dpp/TGF beta signaling pathway and encodes a transcription factor homologous to the human MBP family. *Cell*, 81(5), pp.781–790.

Asai-Coakwell, M. et al., 2013. Contribution of growth differentiation factor 6-dependent cell survival to early-onset retinal dystrophies. *Human Molecular Genetics*, 22(7), pp.1432–1442.

Asai-Coakwell, M. et al., 2007. GDF6, a novel locus for a spectrum of ocular developmental anomalies. *Am J Hum Genet*, 80, pp.306 – 315.

Asai-Coakwell, M. et al., 2009. Incomplete penetrance and phenotypic variability characterize Gdf6-attributable oculo-skeletal phenotypes. *Hum Mol Genet*, 18, pp.1110 – 1121.

Bailey, T.J. et al., 2004. Regulation of vertebrate eye development by Rx genes. *The International Journal of Developmental Biology*, 48(8-9), pp.761–770.

Baker, N.E., 2001. Master regulatory genes; telling them what to do. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 23(9), pp.763–766.

Bakrania, P. et al., 2008. Mutations in BMP4 cause eye, brain, and digit developmental anomalies: overlap between the BMP4 and hedgehog signaling pathways. *American Journal of Human Genetics*, 82(2), pp.304–319.

Barkai, N. & Ben-Zvi, D., 2009. 'Big frog, small frog'--maintaining proportions in embryonic development: delivered on 2 July 2008 at the 33rd FEBS Congress in Athens, Greece. *The FEBS journal*, 276(5), pp.1196–1207.

Barro, O. et al., 1995. Widespread expression of the eve1 gene in zebrafish embryos affects the anterior-posterior axis pattern. *Developmental Genetics*, 17(2), pp.117–128.

Bauer, H. et al., 2001. The type I serine/threonine kinase receptor Alk8/Lost-a-fin is required for Bmp2b/7 signal transduction during dorsoventral patterning of the zebrafish embryo. *Development (Cambridge, England)*, 128(6), pp.849–858.

Bedell, V.M. et al., 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature*, 491(7422), pp.114–118.

Bedell, V.M., Westcot, S.E. & Ekker, S.C., 2011. Lessons from morpholino-based screening in zebrafish. *Briefings in Functional Genomics*. Available at: http://bfg.oxfordjournals.org/content/early/2011/07/10/bfgp.elr021.abstract.

Belenkaya, T.Y. et al., 2004. Drosophila Dpp morphogen movement is independent of dynamin-mediated endocytosis but regulated by the glypican members of heparan sulfate proteoglycans. *Cell*, 119(2), pp.231–244.

Belenkaya, T.Y. et al., 2004. Drosophila Dpp morphogen movement is independent of dynamin-mediated endocytosis but regulated by the glypican members of heparan sulfate proteoglycans. *Cell*, 119(2), pp.231–244.

Berget, S.M., Moore, C. & Sharp, P.A., 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(8), pp.3171–3175.

Bielen, H. & Houart, C., 2012. BMP signaling protects telencephalic fate by repressing eye identity and its Cxcr4-dependent morphogenesis. *Developmental Cell*, 23(4), pp.812–822.

Bill, B.R. et al., 2009. A primer for morpholino use in zebrafish. *Zebrafish*, 6(1), pp.69–77.

Blackburn, P.R. et al., 2013. The CRISPR system--keeping zebrafish gene targeting fresh. *Zebrafish*, 10(1), pp.116–118.

Blitz, I.L. & Cho, K.W.Y., 2009. Finding partners: how BMPs select their targets. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 238(6), pp.1321–1331.

Boch, J. et al., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, 326(5959), pp.1509–1512.

Bogdanove, A.J. & Voytas, D.F., 2011. TAL effectors: customizable proteins for DNA targeting. *Science (New York, N.Y.)*, 333(6051), pp.1843–1846.

Branam, A.M. et al., 2010. Zebrafish chordin-like and chordin are functionally redundant in regulating patterning of the dorsoventral axis. *Developmental biology*, 341(2), pp.444–458.

Brenet, F. et al., 2011. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PloS one*, 6(1), p.e14524.

Briggs, A.W. et al., 2012. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic acids research*, 40(15), p.e117.

Brown, K.E. et al., 2010. Nlcam modulates midline convergence during anterior neural plate morphogenesis. *Developmental Biology*, 339(1), pp.14–25.

Brownlie, A. et al., 1998. Positional cloning of the zebrafish sauternes gene: a model for congenital sideroblastic anaemia. *Nature genetics*, 20(3), pp.244–250.

Brown TA., 2002. *Genomes.* 2nd edition., Oxford: Wiley-Liss.

Bultmann, S. et al., 2012. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic acids research*, 40(12), pp.5368–5377.

Cade, L. et al., 2012. Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic acids research*, 40(16), pp.8001–8010.

Cermak, T. et al., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, 39(12), p.e82.

Certo, M.T. et al., 2012. Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nature methods*, 9(10), pp.973–975.

Chang, C. et al., 2001. Twisted gastrulation can function as a BMP antagonist. *Nature*, 410(6827), pp.483–487.

Chang, N. et al., 2013. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell research*, 23(4), pp.465–472.

Chassaing, N. et al., 2014. Molecular findings and clinical data in a cohort of 150 patients with anophthalmia/microphthalmia. *Clinical Genetics*, 86(4), pp.326–334.

Chen, F.-C. et al., 2006. Identification and evolutionary analysis of novel exons and alternative splicing events using cross-species EST-to-genome comparisons in human, mouse and rat. *BMC bioinformatics*, 7, p.136.

Cheng, L.E. & Reed, R.R., 2007. Zfp423/OAZ participates in a developmental switch during olfactory neurogenesis. *Neuron*, 54(4), pp.547–557.

Cheng, L.E., Zhang, J. & Reed, R.R., 2007. The transcription factor Zfp423/OAZ is required for cerebellar development and CNS midline patterning. *Developmental Biology*, 307(1), pp.43–52.

Chen, S. et al., 2013. A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic acids research*, 41(4), pp.2769–2778.

Chiang, C. et al., 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature*, 383(6599), pp.407–413.

Chitayat, D. et al., 2007. The PDAC syndrome (pulmonary hypoplasia/agenesis, diaphragmatic hernia/eventration, anophthalmia/microphthalmia, and cardiac defect) (Spear syndrome, Matthew-Wood syndrome): report of eight cases

including a living child and further evidence for autosomal recessive inheritance. *American Journal of Medical Genetics. Part A*, 143A(12), pp.1268–1281.

Chow, L.T. et al., 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12(1), pp.1–8.

Chow, R.L. & Lang, R.A., 2001. Early eye development in vertebrates. *Annual Review of Cell and Developmental Biology*, 17, pp.255–296.

Christian, M.L. et al., 2012. Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PloS one*, 7(9), p.e45383.

Chuang, J.C. & Raymond, P.A., 2002. Embryonic origin of the eyes in teleost fish. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 24(6), pp.519–529.

Clark, K.J., Voytas, D.F. & Ekker, S.C., 2011. A TALE of two nucleases: gene targeting for the masses? *Zebrafish*, 8(3), pp.147–149.

Cogulu, O. et al., 2000. Waardenburg anophthalmia syndrome: report and review. *American Journal of Medical Genetics*, 90(2), pp.173–174.

Cohen, M.M., 2009. Perspectives on RUNX genes: an update. *American Journal of Medical Genetics. Part A*, 149A(12), pp.2629–2646.

Collery, R.F. & Link, B.A., 2011. Dynamic smad-mediated BMP signaling revealed through transgenic zebrafish. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 240(3), pp.712–722.

Cong, L. et al., 2012. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun*, 3, p.968.

Da Costa, M.M.J. et al., 2014. A new zebrafish model produced by TILLING of SOD1-related amyotrophic lateral sclerosis replicates key features of the disease and represents a tool for in vivo therapeutic screening. *Disease Models & Mechanisms*, 7(1), pp.73–81.

Dahlem, T.J. et al., 2012. Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS genetics*, 8(8), p.e1002861.

Dai, H. et al., 2000. The zinc finger protein schnurri acts as a Smad partner in mediating the transcriptional response to decapentaplegic. *Developmental Biology*, 227(2), pp.373–387.

Danno, H. et al., 2008. Molecular links among the causative genes for ocular malformation: Otx2 and Sox2 coregulate Rax expression. *Proceedings of the National Academy of Sciences of the United States of America*, 105(14), pp.5408–5413.

Das, S. et al., 1995. Similar splicing mutations of the Menkes/mottled copper-transporting ATPase gene in occipital horn syndrome and the blotchy mouse. *American Journal of Human Genetics*, 56(3), pp.570–576.

Dattani, M.T. et al., 1998. Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. *Nature Genetics*, 19(2), pp.125–133.

David Sexton, 2010. *Smoc-1 an antagonist of BMP signalling.* University of Edinburgh.

Delot, E. et al., 1999. The BMP-related protein radar: a maintenance factor for dorsal neuroectoderm cells? *Mech Dev*, 85, pp.15 – 25.

Deng, D. et al., 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science (New York, N.Y.)*, 335(6069), pp.720–723.

Dick, A., Meier, A. & Hammerschmidt, M., 1999. Smad1 and Smad5 have distinct roles during dorsoventral patterning of the zebrafish embryo. *Developmental dynamics: an official publication of the American Association of Anatomists*, 216(3), pp.285–298.

Ding, Q. et al., 2013. A TALEN genome-editing system for generating human stem cell-based disease models. *Cell stem cell*, 12(2), pp.238–251.

Donner, A.L., Lachke, S.A. & Maas, R.L., 2006. Lens induction in vertebrates: variations on a conserved theme of signaling events. *Seminars in Cell & Developmental Biology*, 17(6), pp.676–685.

Dorsky, R.I., Sheldahl, L.C. & Moon, R.T., 2002. A transgenic Lef1/beta-catenin-dependent reporter is expressed in spatially restricted domains throughout zebrafish development. *Developmental Biology*, 241(2), pp.229–237.

Doyle, E.L. et al., 2012. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic acids research*, 40(Web Server issue), pp.W117–122.

Doyon, Y. et al., 2011. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature methods*, 8(1), pp.74–79.

Dreieicher, E. et al., 2009. Nitric Oxide Inhibits Glomerular TGF-{beta} Signaling via SMOC-1. *J Am Soc Nephrol*, 20(9), pp.1963–1974.

Driever, W. et al., 1996. A genetic screen for mutations affecting embryogenesis in zebrafish. *Development*, 123, pp.37 – 46.

Dudley, A.T., Lyons, K.M. & Robertson, E.J., 1995. A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. *Genes & Development*, 9(22), pp.2795–2807.

Duester, G., 2008. Retinoic acid synthesis and signaling during early organogenesis. *Cell*, 134(6), pp.921–931.

Dunker, A.K. et al., 2008. Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology*, 18(6), pp.756–764.

Dyson, H.J. & Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. *Nature Reviews. Molecular Cell Biology*, 6(3), pp.197–208.

Ehrlich, M. et al., 2011. Homomeric and heteromeric complexes among TGF-β and BMP receptors and their roles in signaling. *Cellular Signalling*, 23(9), pp.1424–1432.

Eichler, E.E., 2001. Segmental duplications: what's missing, misassigned, and misassembled--and should we care? *Genome Research*, 11(5), pp.653–656.

Eisen, J. & Smith, J., 2008. Controlling morpholino experiments: don't stop making antisense. *Development*,135(10), pp.1735–1743.

Ekker, S.C., 2008. Zinc finger-based knockout punches for zebrafish genes. *Zebrafish*, 5(2), pp.121–123.

Ekker, S.C. & Larson, J.D., 2001. Morphant technology in model developmental systems. *Genesis*, 30, pp.89 – 93.

Eldar, A. et al., 2003. Self-enhanced ligand degradation underlies robustness of morphogen gradients. *Developmental Cell*, 5(4), pp.635–646.

Engeszer, R.E. et al., 2007. Zebrafish in the wild: a review of natural history and new notes from the field. *Zebrafish*, 4(1), pp.21–40.

Engler, C. et al., 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PloS one*, 4(5), p.e5553.

Engler, C., Kandzia, R. & Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. *PloS one*, 3(11), p.e3647.

Entchev, E.V., Schwabedissen, A. & González-Gaitán, M., 2000. Gradient formation of the TGF-beta homolog Dpp. *Cell*, 103(6), pp.981–991.

Eudy, J.D. et al., 1998. Mutation of a gene encoding a protein with extracellular matrix motifs in Usher syndrome type IIa. *Science (New York, N.Y.)*, 280(5370), pp.1753–1757.

Fantes, J. et al., 2003. Mutations in SOX2 cause anophthalmia. *Nature Genetics*, 33(4), pp.461–463.

Fares-Taie, L. et al., 2013. ALDH1A3 mutations cause recessive anophthalmia and microphthalmia. *American Journal of Human Genetics*, 92(2), pp.265–270.

Fay, D.S., 2005. The cell cycle and development: lessons from C. elegans. *Seminars in Cell & Developmental Biology*, 16(3), pp.397–406.

Fernández-Garre, P. et al., 2002. Fate map of the chicken neural plate at stage 4. *Development (Cambridge, England)*, 129(12), pp.2807–2822.

Finckbeiner, S. et al., 2011. Transient knockdown and overexpression reveal a developmental role for the zebrafish enosf1b gene. *Cell & bioscience*, 1, p.32.

Fitzpatrick, D.R. & van Heyningen, V., 2005. Developmental eye disorders. *Current Opinion in Genetics & Development*, 15(3), pp.348–353.

French, C.R. et al., 2009. Gdf6a is required for the initiation of dorsal-ventral retinal patterning and lens development. *Developmental Biology*, 333(1), pp.37–47.

Fujii, M. et al., 1999. Roles of bone morphogenetic protein type I receptors and Smad proteins in osteoblast and chondroblast differentiation. *Molecular biology of the cell*, 10(11), pp.3801–3813.

Fürthauer, M. et al., 2004. Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. *Development (Cambridge, England)*, 131(12), pp.2853–2864.

Furuta, Y. & Hogan, B.L.M., 1998. BMP4 is essential for lens induction in the mouse embryo. *Genes & Development*, 12(23), pp.3764–3775.

Fu, Y. et al., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, 31(9), pp.822–826.

Garavelli, L. et al., 2006. Anophthalmos with limb anomalies (Waardenburg opththalmo-acromelic syndrome): report of a new Italian case with renal anomaly and review. *Genetic Counseling (Geneva, Switzerland)*, 17(4), pp.449–455.

Gerety, S.S. & Wilkinson, D.G., 2011. Morpholino artifacts provide pitfalls and reveal a novel role for pro-apoptotic genes in hindbrain boundary development. *Developmental biology*, 350(2), pp.279–289.

Gersdorff, N. et al., 2006. Secreted modular calcium-binding protein-1 localization during mouse embryogenesis. *Histochemistry and Cell Biology*, 126(6), pp.705–712.

Gerth-Kahlert, C. et al., 2013. Clinical and mutation analysis of 51 probands with anophthalmia and/or severe microphthalmia from a single center. *Molecular Genetics & Genomic Medicine*, 1(1), pp.15–31.

Glaser, T. et al., 1994. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nature Genetics*, 7(4), pp.463–471.

Glass, A.S. & Dahm, R., 2004. The zebrafish as a model organism for eye development. *Ophthalmic Research*, 36(1), pp.4–24.

Gong, C. et al., 2005. Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nature Structural & Molecular Biology*, 12(4), pp.304–312.

Gosse, N.J. & Baier, H., 2009. An essential role for Radar (Gdf6a) in inducing dorsal fate in the zebrafish retina. *Proc Natl Acad Sci USA*, 106, pp.2236 – 2241.

Gritsman, K. et al., 1999. The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. *Cell*, 97(1), pp.121–132.

Grotewold, L. et al., 2001. Bambi is coexpressed with Bmp-4 during mouse embryogenesis. *Mechanisms of development*, 100(2), pp.327–330.

Di Guglielmo, G.M. et al., 2003. Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. *Nature Cell Biology*, 5(5), pp.410–421.

Haffter, P. et al., 1996. The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development*, 123, pp.1 – 36.

Halder, G., Callaerts, P. & Gehring, W.J., 1995. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science (New York, N.Y.)*, 267(5205), pp.1788–1792.

Hamaratoglu, F. et al., 2011. Dpp signaling activity requires Pentagone to scale with tissue size in the growing Drosophila wing imaginal disc. *PLoS biology*, 9(10), p.e1001182.

Hammerschmidt, M. & Mullins, M.C., 2002. Dorsoventral patterning in the zebrafish: bone morphogenetic proteins and beyond. *Results and Problems in Cell Differentiation*, 40, pp.72–95.

Hanel, M.L. & Hensey, C., 2006. Eye and neural defects associated with loss of GDF6. *BMC Dev Biol*, 6, p.43.

Von der Hardt, S. et al., 2007. The Bmp gradient of the zebrafish gastrula guides migrating lateral cells by regulating cell-cell adhesion. *Current Biology: CB*, 17(6), pp.475–487.

Hartung, A. et al., 2006. Different routes of bone morphogenic protein (BMP) receptor endocytosis influence BMP signaling. *Molecular and Cellular Biology*, 26(20), pp.7791–7805.

Hata, A. et al., 1998. Smad6 inhibits BMP/Smad1 signaling by specifically competing with the Smad4 tumor suppressor. *Genes & Development*, 12(2), pp.186–197.

Hayashi, H. et al., 1997. The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. *Cell*, 89(7), pp.1165–1173.

Von Heijne, G. & Gavel, Y., 1988. Topogenic signals in integral membrane proteins. *European journal of biochemistry / FEBS*, 174(4), pp.671–678.

Heisenberg, C.P. et al., 2001. A mutation in the Gsk3-binding domain of zebrafish Masterblind/Axin1 leads to a fate transformation of telencephalon and eyes to diencephalon. *Genes & Development*, 15(11), pp.1427–1434.

Hellsten, U. et al., 2010. The genome of the Western clawed frog Xenopus tropicalis. *Science (New York, N.Y.)*, 328(5978), pp.633–636.

Hild, M. et al., 1999. The smad5 mutation somitabun blocks Bmp2b signaling during early dorsoventral patterning of the zebrafish embryo. *Development (Cambridge, England)*, 126(10), pp.2149–2159.

Hingorani, M., Hanson, I. & van Heyningen, V., 2012. Aniridia. *European journal of human genetics: EJHG*, 20(10), pp.1011–1017.

Houart, C. et al., 2002. Establishment of the telencephalon during gastrulation by local antagonism of Wnt signaling. *Neuron*, 35(2), pp.255–265.

Howe, K. et al., 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446), pp.498–503.

Hinaux H. et al., 2011. A developmental staging table for Astyanax mexicanus surface fish and Pachón cavefish. *Zebrafish*, 8, pp.155–165.

Hruscha, A. et al., 2013. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development (Cambridge, England)*, 140(24), pp.4982–4987.

Hsu, L.C. et al., 1994. Molecular cloning, genomic organization, and chromosomal localization of an additional human aldehyde dehydrogenase gene, ALDH6. *Genomics*, 24(2), pp.333–341.

Huang, P. et al., 2012. Reverse genetic approaches in zebrafish. *Journal of genetics and genomics = Yi chuan xue bao*, 39(9), pp.421–433.

Hu, M.C. et al., 2004. p38MAPK acts in the BMP7-dependent stimulatory pathway during epithelial cell morphogenesis and is regulated by Smad1. *The Journal of Biological Chemistry*, 279(13), pp.12051–12059.

Hwang, W.Y. et al., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology*, 31(3), pp.227–229.

Hwang, W.Y. et al., 2013. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PloS One*, 8(7), p.e68708.

Imamura, T. et al., 1997. Smad6 inhibits signalling by the TGF-beta superfamily. *Nature*, 389(6651), pp.622–626.

Van Impel, A. et al., 2014. Divergence of zebrafish and mouse lymphatic cell fate specification pathways. *Development (Cambridge, England)*, 141(6), pp.1228–1238.

Inoue, T., Nakamura, S. & Osumi, N., 2000. Fate mapping of the mouse prosencephalic neural plate. *Developmental Biology*, 219(2), pp.373–383.

Jao, L.-E., Wente, S.R. & Chen, W., 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences of the United States of America*, 110(34), pp.13904–13909.

Jinek, M. et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096), pp.816–821.

Kabashi, E. et al., 2011. Zebrafish models for the functional genomics of neurogenetic disorders. *Biochimica et biophysica acta*, 1812(3), pp.335–345.

Kamachi, Y. et al., 2001. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes & Development*, 15(10), pp.1272–1286.

Kara, F. et al., 2002. A case report of prenatally diagnosed ophthalmo-acromelic syndrome type Waardenburg. *Prenatal Diagnosis*, 22(5), pp.395–397.

Karsenty, G., 2008. Transcriptional control of skeletogenesis. *Annual Review of Genomics and Human Genetics*, 9, pp.183–196.

Kataoka, H. et al., 2000. Cloning and embryonic expression patterns of the zebrafish Runt domain genes, runxa and runxb. *Mechanisms of Development*, 98(1-2), pp.139–143.

Kavsak, P. et al., 2000. Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. *Molecular Cell*, 6(6), pp.1365–1375.

Kawaguchi, R. et al., 2007. A membrane receptor for retinol binding protein mediates cellular uptake of vitamin A. *Science (New York, N.Y.)*, 315(5813), pp.820–825.

Kelemen, O. et al., 2013. Function of alternative splicing. *Gene*, 514(1), pp.1–30.

Kicheva, A. & González-Gaitán, M., 2008. The Decapentaplegic morphogen gradient: a precise definition. *Current Opinion in Cell Biology*, 20(2), pp.137–143.

Kimble, J. & Hirsh, D., 1979. The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. *Developmental Biology*, 70(2), pp.396–417.

Kim, H. & Kim, J.-S., 2014. A guide to genome engineering with programmable nucleases. *Nature Reviews. Genetics*, 15(5), pp.321–334.

Kimmel, C.B. et al., 1995. Stages of Embryonic-Development of the Zebrafish 1. *Developmental Dynamics*, 203, pp.253 – 310.

Kim, Y. et al., 2013. A library of TAL effector nucleases spanning the human genome. *Nature Biotechnology*, 31(3), pp.251–258.

Kim, Y.G., Cha, J. & Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), pp.1156–1160.

Kishimoto, Y. et al., 1997. The molecular nature of zebrafish swirl: BMP2 function is essential during early dorsoventral patterning. *Development (Cambridge, England)*, 124(22), pp.4457–4466.

Klemenčič, M. et al., 2013. The heparin-binding activity of secreted modular calcium-binding protein 1 (SMOC-1) modulates its cell adhesion properties. *PloS One*, 8(2), p.e56839.

Kondo, M., 2007. Bone morphogenetic proteins in the early development of zebrafish. *The FEBS Journal*, 274(12), pp.2960–2967.

Koos, D.S. & Ho, R.K., 1999. The nieuwkoid/dharma homeobox gene is essential for bmp2b repression in the zebrafish pregastrula. *Developmental biology*, 215(2), pp.190–207.

Koressaar, T. & Remm, M., 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics (Oxford, England)*, 23(10), pp.1289–1291.

Koshiba-Takeuchi, K. et al., 2000. Tbx5 and the retinotectum projection. *Science (New York, N.Y.)*, 287(5450), pp.134–137.

Krawczak, M. et al., 2007. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Human Mutation*, 28(2), pp.150–158.

Kriventseva, E.V. et al., 2003. Increase of functional diversity by alternative splicing. *Trends in genetics: TIG*, 19(3), pp.124–128.

Kruse-Bend, R. et al., 2012. Extraocular ectoderm triggers dorsal retinal fate during optic vesicle evagination in zebrafish. *Developmental Biology*, 371(1), pp.57–65.

Lachke, S.A. & Maas, R.L., 2010. Building the developmental oculome: systems biology in vertebrate eye development and disease. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, 2(3), pp.305–323.

Lane, B.M. & Lister, J.A., 2012. Otx but not Mitf transcription factors are required for zebrafish retinal pigment epithelium development. *PloS One*, 7(11), p.e49357.

Lapraz, F., Besnardeau, L. & Lepage, T., 2009. Patterning of the dorsal-ventral axis in echinoderms: insights into the evolution of the BMP-chordin signaling network. *PLoS biology*, 7(11), p.e1000248.

Le, H.-G.T., Dowling, J.E. & Cameron, D.J., 2012. Early retinoic acid deprivation in developing zebrafish results in microphthalmia. *Visual Neuroscience*, 29(4-5), pp.219–228.

Lele, Z., Bakkers, J. & Hammerschmidt, M., 2001. Morpholino phenocopies of the swirl, snailhouse, somitabun, minifin, silberblick, and pipetail mutations. *Genesis (New York, N.Y.: 2000)*, 30(3), pp.190–194.

Lepage, S.E. & Bruce, A.E.E., 2010. Zebrafish epiboly: mechanics and mechanisms. *The International Journal of Developmental Biology*, 54(8-9), pp.1213–1228.

Lewis, B.P., Green, R.E. & Brenner, S.E., 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences*, 100(1), pp.189–192.

Li et al., 1999. Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome informatics. Workshop on Genome Informatics*, 10, pp.30–40.

Li, S. & Koromilas, A.E., 2001. Dominant negative function by an alternatively spliced form of the interferon-inducible protein kinase PKR. *The Journal of biological chemistry*, 276(17), pp.13881–13890.

Little, S.C. & Mullins, M.C., 2009. Bone morphogenetic protein heterodimers assemble heteromeric type I receptor complexes to pattern the dorsoventral axis. *Nat Cell Biol*, 11(5), pp.637–643.

Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, 25(4), pp.402–408.

Lu, J. et al., 2010. Alternative splicing in teleost fish genomes: same-species and cross-species analysis and comparisons. *Molecular genetics and genomics: MGG*, 283(6), pp.531–539.

Macdonald, R. et al., 1995. Midline signalling is required for Pax gene regulation and patterning of the eyes. *Development (Cambridge, England)*, 121(10), pp.3267–3278.

Macdonald, R. & Wilson, S.W., 1997. Distribution of Pax6 protein during eye development suggests discrete roles in proliferative and differentiated visual cells. *Dev Genes Evol*, 206, pp.363 – 369.

Maeno, M. et al., 1996. The role of BMP-4 and GATA-2 in the induction and differentiation of hematopoietic mesoderm in Xenopus laevis. *Blood*, 88(6), pp.1965–1972.

Magen, A. & Ast, G., 2005. The importance of being divisible by three in alternative splicing. *Nucleic acids research*, 33(17), pp.5574–5582.

Mak, A.N.-S. et al., 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science (New York, N.Y.)*, 335(6069), pp.716–719.

Mali, P. et al., 2013. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), pp.823–826.

Mani, M. et al., 2005. Binding of two zinc finger nuclease monomers to two specific sites is required for effective double-strand DNA cleavage. *Biochemical and biophysical research communications*, 334(4), pp.1191–1197.

Manzini, M.C. et al., 2012. Exome sequencing and functional validation in zebrafish identify GTDC2 mutations as a cause of Walker-Warburg syndrome. *American journal of human genetics*, 91(3), pp.541–547.

Marsh-Armstrong, N. et al., 1994. Retinoic acid is necessary for development of the ventral retina in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America*, 91(15), pp.7286–7290.

Mart�nez-Barber�, J.P. et al., 1997. Cloning and expression of three members of the zebrafish Bmp family: Bmp2a, Bmp2b and Bmp4. *Gene*, 198(1-2), pp.53–59.

Martinez-Morales, J.R. et al., 2001. Otx genes are required for tissue specification in the developing eye. *Development (Cambridge, England)*, 128(11), pp.2019–2030.

Mashimo, T. et al., 2013. Efficient gene targeting by TAL effector nucleases coinjected with exonucleases in zygotes. *Sci. Rep.*, 3. Available at: http://dx.doi.org/10.1038/srep01253.

Mathers, P.H. et al., 1997. The Rx homeobox gene is essential for vertebrate eye development. *Nature*, 387(6633), pp.603–607.

Matsuo, I. et al., 1995. Mouse Otx2 functions in the formation and patterning of rostral head. *Genes & Development*, 9(21), pp.2646–2658.

McCurley, A.T. & Callard, G.V., 2008. Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. *BMC molecular biology*, 9, p.102.

McGary, K.L. et al., 2010. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proceedings of the National Academy of Sciences*, 107(14), pp.6544 –6549.

Meckler, J.F. et al., 2013. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic acids research*, 41(7), pp.4118–4128.

Merriman, B., Ion Torrent R&D Team & Rothberg, J.M., 2012. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23), pp.3397–3417.

Meyer, A. & Schartl, M., 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current opinion in cell biology*, 11(6), pp.699–704.

Miller, J.C. et al., 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature biotechnology*, 25(7), pp.778–785.

Miller, J.C. et al., 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotech*, 29(2), pp.143–148.

Mitchell, H. et al., 2004. Ligand-dependent and -independent transforming growth factor-beta receptor recycling regulated by clathrin-mediated endocytosis and Rab11. *Molecular Biology of the Cell*, 15(9), pp.4166–4178.

Miyazono, K., Kamiya, Y. & Morikawa, M., 2010. Bone morphogenetic protein receptors and signal transduction. *Journal of Biochemistry*, 147(1), pp.35–51.

Mladenov, E. & Iliakis, G., 2011. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutation research*, 711(1-2), pp.61–72.

Modrek, B. & Lee, C.J., 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature Genetics*, 34(2), pp.177–180.

Moens, C.B. et al., 2008. Reverse genetics in zebrafish by TILLING. *Briefings in functional genomics & proteomics*, 7(6), pp.454–459.

Molina, G.A., Watkins, S.C. & Tsang, M., 2007. Generation of FGF reporter transgenic zebrafish and their utility in chemical screens. *BMC developmental biology*, 7, p.62.

Møller, L.B. et al., 2000. Similar splice-site mutations of the ATP7A gene lead to different phenotypes: classical Menkes disease or occipital horn syndrome. *American Journal of Human Genetics*, 66(4), pp.1211–1220.

Monte Westerfield, 2000. *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio).*, Univ. of Oregon Press, Eugene.: Univ. of Oregon Press, Eugene.

Moore, F.E. et al., 2012. Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). *PloS one*, 7(5), p.e37877.

Morcos, P.A., 2007. Achieving targeted and quantifiable alteration of mRNA splicing with Morpholino oligos. *Biochemical and Biophysical Research Communications*, 358(2), pp.521–527.

Moscou, M.J. & Bogdanove, A.J., 2009. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science*, 326(5959), pp.1501–1501.

Mullins, M.C. et al., 1996. Genes establishing dorsoventral pattern formation in the zebrafish embryo: the ventral specifying genes. *Development (Cambridge, England)*, 123, pp.81–93.

Muntoni, F., Torelli, S. & Ferlini, A., 2003. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurology*, 2(12), pp.731–740.

Mussolino, C. et al., 2014. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic acids research*, 42(10), pp.6762–6773.

Nasevicius, A. & Ekker, S.C., 2000. Effective targeted gene 'knockdown' in zebrafish. *Nature Genetics*, 26(2), pp.216–220.

Navaneetha Krishnan Bharathan, 2014. *The role of the rx3/ otx pathway in zebrafish eye developemnt*. Master of Science. Virginia Commonwealth University.

Nieuwkoop, P.D., 1963. Pattern formation in artificially activated ectoderm (Rana pipiens and Ambystoma punctatum). *Developmental Biology*, 6, pp.255–279.

Nikaido, M. et al., 1999. In vivo analysis using variants of zebrafish BMPR-IA: range of action and involvement of BMP in ectoderm patterning. *Development (Cambridge, England)*, 126(1), pp.181–190.

Nusslein-Volhard, C. & Dahm, R., 2002. Zebrafish: A Practical Approach. In London: Oxford University Press, pp. 39–44.

Ohkawara, B. et al., 2002. Action range of BMP is defined by its N-terminal basic amino acid core. *Current Biology: CB*, 12(3), pp.205–209.

Okada, I. et al., 2011. SMOC1 is essential for ocular and limb development in humans and mice. *American Journal of Human Genetics*, 88(1), pp.30–41.

Oren, T., Torregroza, I. & Evans, T., 2005. An Oct-1 binding site mediates activation of the gata2 promoter by BMP signaling. *Nucleic Acids Research*, 33(13), pp.4357–4367.

Ota, S. et al., 2014. Multiple genome modifications by the CRISPR/Cas9 system in zebrafish. *Genes to cells: devoted to molecular & cellular mechanisms*.

Pancsa, R. & Fuxreiter, M., 2012. Interactions via intrinsically disordered regions: what kind of motifs? *IUBMB life*, 64(6), pp.513–520.

Pan, Q. et al., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40(12), pp.1413–1415.

Pappano, W.N. et al., 2003. Use of Bmp1/Tll1 doubly homozygous null mice and proteomics to identify and validate in vivo substrates of bone morphogenetic protein 1/tolloid-like metalloproteinases. *Molecular and Cellular Biology*, 23(13), pp.4428–4438.

Paradis, E., Claude, J. & Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics (Oxford, England)*, 20(2), pp.289–290.

Pasutto, F. et al., 2007. Mutations in STRA6 cause a broad spectrum of malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation. *American Journal of Human Genetics*, 80(3), pp.550–560.

Pauls, S., Geldmacher-Voss, B. & Campos-Ortega, J.A., 2001. A zebrafish histone variant H2A.F/Z and a transgenic H2A.F/Z:GFP fusion protein for in vivo studies of embryonic development. *Development Genes and Evolution*, 211(12), pp.603–610.

Peter D. Burge, 2009. *Children's Orthopaedics and Fractures*, Churchill Livingstone.

Petersen, T.N. et al., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*, 8(10), pp.785–786.

Peterson, R.T. et al., 2004. Chemical suppression of a genetic mutation in a zebrafish model of aortic coarctation. *Nature biotechnology*, 22(5), pp.595–599.

Piccolo, S. et al., 1996. Dorsoventral patterning in Xenopus: inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell*, 86(4), pp.589–598.

Pires-daSilva, A. & Sommer, R.J., 2003. The evolution of signalling pathways in animal development. *Nature Reviews. Genetics*, 4(1), pp.39–49.

Plouhinec, J.-L. & De Robertis, E.M., 2009. Systems biology of the self-regulating morphogenetic gradient of the Xenopus gastrula. *Cold Spring Harbor Perspectives in Biology*, 1(2), p.a001701.

Prou, D. et al., 2001. Intracellular retention of the two isoforms of the D(2) dopamine receptor promotes endoplasmic reticulum disruption. *Journal of cell science*, 114(Pt 19), pp.3517–3527.

Pyati, U.J., Webb, A.E. & Kimelman, D., 2005. Transgenic zebrafish reveal stage-specific roles for Bmp signaling in ventral and posterior mesoderm development. *Development (Cambridge, England)*, 132(10), pp.2333–2343.

Pyrowolakis, G. et al., 2004. A simple molecular complex mediates widespread BMP-induced repression during Drosophila development. *Developmental Cell*, 7(2), pp.229–240.

Ragge, N.K. et al., 2005. Heterozygous mutations of OTX2 cause severe ocular malformations. *American Journal of Human Genetics*, 76(6), pp.1008–1022.

Rainger, J. et al., 2011. Loss of the BMP Antagonist, SMOC-1, Causes Ophthalmo-Acromelic (Waardenburg Anophthalmia) Syndrome in Humans and Mice. *PLoS Genetics*, 7(7), p.e1002114.

Rainger, J. et al., 2014. Monoallelic and biallelic mutations in MAB21L2 cause a spectrum of major eye malformations. *American Journal of Human Genetics*, 94(6), pp.915–923.

Ramel, M.-C. & Hill, C.S., 2012. Spatial regulation of BMP activity. *FEBS letters*, 586(14), pp.1929–1941.

Ramirez, C.L. et al., 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Meth*, 5(5), pp.374–375.

Ran, F.A. et al., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6), pp.1380–1389.

Reis, L.M. et al., 2011. BMP4 loss-of-function mutations in developmental eye disorders including SHORT syndrome. *Human Genetics*, 130(4), pp.495–504.

Rembold, M. et al., 2006. Individual cell migration serves as the driving force for optic vesicle evagination. *Science (New York, N.Y.)*, 313(5790), pp.1130–1134.Rétaux S. & Casane D., 2013. Evolution of eye development in the darkness of caves: adaption, drift, or both? *Evodevo*, 431, pp. 844–847.

Reyon, D. et al., 2013. Engineering customized TALE nucleases (TALENs) and TALE transcription factors by fast ligation-based automatable solid-phase high-throughput (FLASH) assembly. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, Chapter 12, p.Unit 12.16.

Reyon, D. et al., 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology*, 30(5), pp.460–465.

Rhinn, M. & Dollé, P., 2012. Retinoic acid signalling during development. *Development (Cambridge, England)*, 139(5), pp.843–858.

Robu, M.E. et al., 2007. p53 activation by knockdown technologies. *PLoS genetics*, 3(5), p.e78.

Romero, Obradovic & Dunker, 1997. Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome informatics. Workshop on Genome Informatics*, 8, pp.110–124.

Romero, P. et al., 2001. Sequence complexity of disordered protein. *Proteins*, 42(1), pp.38–48.

Romero, P.R. et al., 2006. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proceedings of the National Academy of Sciences of the United States of America*, 103(22), pp.8390–8395.

Roy, A. et al., 2013. LHX2 is necessary for the maintenance of optic identity and for the progression of optic morphogenesis. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(16), pp.6877–6884.

Riesch R. et al., 2011. Shared and unique patterns of embryo development in extremophile poeciliids. *PLoS One*, 6(11), e27377

Sakurai, Y. et al., 2001. Novel protein kinase C delta isoform insensitive to caspase-3. *Biological & pharmaceutical bulletin*, 24(9), pp.973–977.

Sakuta, H. et al., 2001. Ventroptin: a BMP-4 antagonist expressed in a double-gradient pattern in the retina. *Science (New York, N.Y.)*, 293(5527), pp.111–115.

Sander, J.D., Dahlborg, E.J., et al., 2011. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nature methods*, 8(1), pp.67–69.

Sander, J.D., Cade, L., et al., 2011. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature biotechnology*, 29(8), pp.697–698.

Sasagawa, S. et al., 2002. Axes establishment during eye morphogenesis in Xenopus by coordinate and antagonistic actions of BMP4, Shh, and RA. *Genesis (New York, N.Y.: 2000)*, 33(2), pp.86–96.

Schindelin, J. et al., 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), pp.676–682.

Schulte-Merker, S. et al., 1997. The zebrafish organizer requires chordino. *Nature*, 387(6636), pp.862–863.

Schulte-Merker, S. & Stainier, D.Y.R., 2014. Out with the old, in with the new: reassessing morpholino knockdowns in light of genome editing technology. *Development (Cambridge, England)*, 141(16), pp.3103–3104.

Schwank, G. et al., 2011. Formation of the long range Dpp morphogen gradient. *PLoS biology*, 9(7), p.e1001111.

Seebald, J.L. & Szeto, D.P., 2011. Zebrafish eve1 regulates the lateral and ventral fates of mesodermal progenitor cells at the onset of gastrulation. *Developmental Biology*, 349(1), pp.78–89.

Sharpe, J. et al., 2002. Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science (New York, N.Y.)*, 296(5567), pp.541–545.

Shimizu, N., Kawakami, K. & Ishitani, T., 2012. Visualization and exploration of Tcf/Lef function using a highly responsive Wnt/β-catenin signaling-reporter transgenic zebrafish. *Developmental Biology*, 370(1), pp.71–85.

Sieber, C. et al., 2009. Recent advances in BMP receptor signaling. *Cytokine & Growth Factor Reviews*, 20(5-6), pp.343–355.

Sivasubbu, S., Sachidanandan, C. & Scaria, V., 2013. Time for the zebrafish ENCODE. *Journal of genetics*, 92(3), pp.695–701.

Skromne, I. & Prince, V.E., 2008a. Current perspectives in zebrafish reverse genetics: Moving forward. *Developmental Dynamics*, 237(4), pp.861–882.

Skromne, I. & Prince, V.E., 2008b. Current perspectives in zebrafish reverse genetics: Moving forward. *Developmental Dynamics*, 237(4), pp.861–882.

S Olitsky and L Nelson, 2006. *Duane's clinical ophthalmology*, Williams & Wilkins. Available at: http://www.ncbi.nlm.nih.gov.ezproxy.is.ed.ac.uk/pubmed/?term=Duane%27s+clinical+ophthalmology. [Accessed October 9, 2014].

Solnica-Krezel, L., 2006. Gastrulation in zebrafish -- all just about adhesion? *Current Opinion in Genetics & Development*, 16(4), pp.433–441.

Solnica-Krezel, L. & Sepich, D.S., 2012. Gastrulation: making and shaping germ layers. *Annual Review of Cell and Developmental Biology*, 28, pp.687–717.

Spence, R. et al., 2008. The behaviour and ecology of the zebrafish, Danio rerio. *Biological reviews of the Cambridge Philosophical Society*, 83(1), pp.13–34.

Srour, M. et al., 2013. Recessive and dominant mutations in retinoic acid receptor beta in cases with microphthalmia and diaphragmatic hernia. *American Journal of Human Genetics*, 93(4), pp.765–772.

Staden, R., 1979. A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*, 6(7), pp.2601–2610.

Stamm, S. et al., 2005. Function of alternative splicing. *Gene*, 344, pp.1–20.

Stemple, D.L., 2004. TILLING--a high-throughput harvest for functional genomics. *Nature reviews. Genetics*, 5(2), pp.145–150.

Streit, A., 2007. The preplacodal region: an ectodermal domain with multipotential progenitors that contribute to sense organs and cranial sensory ganglia. *The International Journal of Developmental Biology*, 51(6-7), pp.447–461.

Sulston, J.E. & Horvitz, H.R., 1977. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Developmental Biology*, 56(1), pp.110–156.

Sun, Ning, 2013. *Engineering of transcription activator-like effector nucleases (TALENs) for targeted genome editing*. PhD. USA: University of Illinois.

Sykes, T.G. et al., 1998. Suppression of GATA factor activity causes axis duplication in Xenopus. *Development (Cambridge, England)*, 125(23), pp.4595–4605.

Tang, R. et al., 2007. Validation of zebrafish (Danio rerio) reference genes for quantitative real-time RT-PCR normalization. *Acta biochimica et biophysica Sinica*, 39(5), pp.384–390.

Taylor, J.S. et al., 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Research*, 13(3), pp.382–390.

Tekin, M., 2002. Ophthalmo-acromelic syndrome. *Orphanet encyclopedia*.

Tekin, M. et al., 2000. Ophthalmo-acromelic syndrome: report and review. *American Journal of Medical Genetics*, 90(2), pp.150–154.

Teleman, A.A. & Cohen, S.M., 2000. Dpp gradient formation in the Drosophila wing imaginal disc. *Cell*, 103(6), pp.971–980.

Thermo Fisher Scientific Inc., 2011. *NanoDrop 1000 Spectrophotometer V3.8 User's Manual*, Thermo Fisher Scientific Inc.

Thomas, J.T. et al., 2009. Xenopus SMOC-1 Inhibits Bone Morphogenetic Protein Signaling Downstream of Receptor Binding and Is Essential for Postgastrulation

Development in Xenopus. *Journal of Biological Chemistry*, 284(28), pp.18994–19005.

Tonkin, M.A., 2009. Failure of differentiation part I: Syndactyly. *Hand Clinics*, 25(2), pp.171–193.

Topczewski, J. et al., 2001. The zebrafish glypican knypek controls cell polarity during gastrulation movements of convergent extension. *Developmental Cell*, 1(2), pp.251–264.

Tsang, M. et al., 2004. A role for MKP3 in axial patterning of the zebrafish embryo. *Development (Cambridge, England)*, 131(12), pp.2769–2779.

Tsang, M. et al., 2000. Zebrafish nma is involved in TGFbeta family signaling. *Genesis (New York, N.Y.: 2000)*, 28(2), pp.47–57.

Untergasser, A. et al., 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Research*, 40(15), p.e115.

Vaillend, C. et al., 2010. Rescue of a dystrophin-like protein by exon skipping in vivo restores GABAA-receptor clustering in the hippocampus of the mdx mouse. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, 18(9), pp.1683–1688.

Vannahme, C. et al., 2002. Characterization of SMOC-1, a Novel Modular Calcium-binding Protein in Basement Membranes. *Journal of Biological Chemistry*, 277(41), pp.37977–37986.

Vannahme, C. et al., 2003. Characterization of SMOC-2, a modular extracellular calcium-binding protein. *Biochem. J.*, 373(3), pp.805–814.

Varga, Z.M., Wegner, J. & Westerfield, M., 1999. Anterior movement of ventral diencephalic precursors separates the primordial eye field in the neural plate and requires cyclops. *Development (Cambridge, England)*, 126(24), pp.5533–5546.

Verheyen, E.M., 2007. Opposing effects of Wnt and MAPK on BMP/Smad signal duration. *Developmental Cell*, 13(6), pp.755–756.

Verma, A.S. & Fitzpatrick, D.R., 2007. Anophthalmia and microphthalmia. *Orphanet J Rare Dis*, 2, p.47.

Voronina, V.A. et al., 2004. Mutations in the human RAX homeobox gene in a patient with anophthalmia and sclerocornea. *Hum Mol Genet*, 13, pp.315 – 322.

Vuilleumier, R. et al., 2010. Control of Dpp morphogen signalling by a secreted feedback regulator. *Nature Cell Biology*, 12(6), pp.611–617.

Wallis, D.E. et al., 1999. Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. *Nature Genetics*, 22(2), pp.196–198.

Wang, D. et al., 2007. Efficient genome-wide mutagenesis of zebrafish genes by retroviral insertions. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30), pp.12428–12433.

Wang, F. et al., 2015. Targeted gene disruption in *Xenopus laevis* using CRISPR/Cas9. *Cell Biosci*, 5(15), pp.1–5.

Wang, H. et al., 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 153(4), pp.910–918.

Wang, P. et al., 2005. Structural genomics analysis of alternative splicing and application to isoform structure modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52), pp.18920–18925.

Ward, A.J. & Cooper, T.A., 2010. The pathobiology of splicing. *The Journal of Pathology*, 220(2), pp.152–163.

Wawersik, S. et al., 1999. BMP7 acts in murine lens placode development. *Developmental Biology*, 207(1), pp.176–188.

Wiedenheft, B., Sternberg, S.H. & Doudna, J.A., 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, 482(7385), pp.331–338.

Van Wijk, E. et al., 2004. Identification of 51 novel exons of the Usher syndrome type 2A (USH2A) gene that encode multiple conserved functional domains and that are mutated in patients with Usher syndrome type II. *American Journal of Human Genetics*, 74(4), pp.738–744.

Wilkinson, R. & Wiedenheft, B., 2014. A CRISPR method for genome engineering. *F1000prime Reports*, 6, p.3.

Williamson, K.A. & FitzPatrick, D.R., 2014. The genetic architecture of microphthalmia, anophthalmia and coloboma. *European Journal of Medical Genetics*.

Wolpert, L., 1969. Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology*, 25(1), pp.1–47.

Woo, K, et al., 1995. Fate maps of the zebrafish embryo. *Curr Opin Genet Dev (Cambridge, England)*, 5, pp. 439–443.

Woo, K. & Fraser, S.E., 1995. Order and coherence in the fate map of the zebrafish nervous system. *Development (Cambridge, England)*, 121(8), pp.2595–2609.

Wyatt, A.W. et al., 2010. Bone morphogenetic protein 7 (BMP7) mutations are associated with variable ocular, brain, ear, palate, and skeletal anomalies. *Human Mutation*, 31(7), pp.781–787.

Xing, Y. & Lee, C.J., 2005. Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. *PLoS genetics*, 1(3), p.e34.

Yamamoto Y. & Jeffery WR., 2000. Central role for the lens in cave fish eye degeneration. *Science*, 289, pp. 631–633.

Yamamoto Y, Stock DW, Jeffery WR., 2004. Hedgehog signalling controls eye degeneration in blind cavefish. *Nature*, 431, pp. 844–847.

Yeh, J.-R.J. et al., 2009. Discovering chemical modifiers of oncogene-regulated hematopoietic differentiation. *Nature chemical biology*, 5(4), pp.236–243.

Ye, M. et al., 2010. Mutation of the bone morphogenetic protein GDF3 causes ocular and skeletal anomalies. *Hum Mol Genet*, 19, pp.287 – 298.

Yu, Y. et al., 2011. Colonization of rice leaf blades by an African strain of Xanthomonas oryzae pv. oryzae depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. *Molecular plant-microbe interactions: MPMI*, 24(9), pp.1102–1113.

Zhang, F. et al., 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology*, 29(2), pp.149–153.

Zhang, L., Mathers, P.H. & Jamrich, M., 2000. Function of Rx, but not Pax6, is essential for the formation of retinal progenitor cells in mice. *Genesis (New York, N.Y.: 2000)*, 28(3-4), pp.135–142.

Zhang, X. et al., 2009. Mutational screening of 10 genes in Chinese patients with microphthalmia and/or coloboma. *Molecular Vision*, 15, pp.2911–2918.

Zuber, M.E. et al., 2003. Specification of the vertebrate eye by a network of eye field transcription factors. *Development (Cambridge, England)*, 130(21), pp.5155–5167.

# Chapter 6

# Materials and Methods

All steps were performed at room temperature unless otherwise stated. If a commercial kit is used accordance with manufacturer instructions unless stated otherwise.

# 6.1 Zebrafish husbandry

The zebrafish husbandry techniques detailed in this section were modified from those found in Westerfield (2000) and Nusslein-Volhard & Dahm (2002 ) which can be consulted if further detail is required.

## 6.1.1 Embryo care

Embryos were obtained via the use of static tanks or using marbles as described in The Zebrafish Book (4th edition Westerfield 2000). Unless otherwise stated the experiments were performed on the AB zebrafish line. Embryos were incubated in 90mm petri dishes (Sterilin) filled with E3 media, at a density of no greater than 100 embryos per dish. All embryos were maintained at 28.5°C. The E3 media the embryos were incubated in was replaced daily and any dead embryos were discarded.

***E3 media***

E3 Media was Prepared in 2 stages following the recipes below.

***E3 Embryo medium 60x Stock solution (1L)***

| | |
|---|---|
| NaCl | 17.2 g |
| KCL | 0.76 g |
| $CaCL_2$ + $2H_20$ | 2.9 g |
| $MgSO_4$ + $7H_2O$ | 4.9 g |
| $ddH_2O$ | make to 1 l |

stored at $4^oC$

***E3 Embryo medium 1x Solution (6L)***

| | |
|---|---|
| 60x E3 Stock solution | 100 ml |
| ddH20 | 5.9 l |
| Methyl Blue | 2 drops |

Stored at rt

# 6.1.2 Embryo Bleaching

To prevent control infections all embryos that were raised beyond 5dpf were leakage bleached which involves sequentially immersing chorionated embryos (up to 2 dpf) in a series of trays using an embryo sieve. Each train is filled with either bleaching solution or washing solution ($ddH_2O$ or E3 medium). The sequence of trays used was as follows:

| Tray | Incubation time |
|---|---|
| 1 | 5 min |
| 2 | 5 min |
| 3 | 5 min |
| 4 | 5 min |
| 5 | 5 min |

The embryos are then transferred into a 90 mm Petri dish and incubated as previously described in **6.1.1**. For transgenic lines the embryos were sorted based on the fluorescence using a Nikon Macroscope. H2A:gfp (Pauls et al. 2001) and rx3:gfp embryos (Rembold et al. 2006) were screened at 1 dpf while BRE:gfp embryos(Collery & Link 2011) were screened at 2 dpf. Embryos that showed weak or no fluorescence were discarded. At 5 dpf the embryos are transferred to the gravity third flowing current aquatic system where they were raised in accordance with the relevant statutes and zebrafish best practice under the care of the fish room technicians.

***bleaching solution (6L)***

| | |
|---|---|
| 10-15% Sodium hypochlorite | 180 µl |
| ddH20 | 500 ml |

Stored at 4°C

## 6.1.3 Dechorionation

Embryos were either dechorionated manually using a pair of watchmaker's forceps and a Nikon SMZ 1000 stereomicroscope or dechorionated chemically using Pronase E.

For chemical Dechorionation 200 µl of Pronase E was added to a 1.5ml Eppendorf containing the embryos suspended in 1 ml of E3 media. The solutions were mixed by inversion and the tubes were incubated for 1 min, before being washed in E3 twice for 2 min each time. After the final wash the embryos were aspirated repeatedly into a plastic graduated pasture pipette. By this point in time almost all of the enzymatically digested chorions will have detached. The embryos are then decanted into a Petri dish and the few remaining attached chorions are removed manually with watchmaker's forceps.

***Pronase E stock solution (12mg/ml)***

| | |
|---|---|
| Pronase E | 12mg |
| E3media | 1ml |

stored at -20$^{o}$C

# 6.1.4 Anaesthesia and Termination

**Anaesthesia**

To anesthetise embryos 1ml of Tricaine solution was added directly to the Petri dishes (~30ml of E3 media). For early stage zebrafish embryos anaesthesia is almost instantaneous.

If embryos were to be retained they were transferred into fresh E3 twice after which time they rapidly recovered and were further incubated.

**Termination**

Adult zebra fish that were old or sick were culled using a standard Schedule 1 procedure. Adult zebra fish were killed with an overdose of anesthetic by immersion in Tricaine solution for at least 10 minutes. The brain was then either destroyed by crushing the head or the fish were decapitated.

***Tricaine Solution(Anaesthesia)***

| | |
|---|---|
| Tricaine powder | 400 mg |
| $ddH_2O$ | 97.9 ml |
| 1M Tris (pH9.0) | ~2.1 ml (Adjust pH to 7) |

***Tricaine Solution(Termination)***

| | |
|---|---|
| Tricaine powder | 400 mg |
| $ddH_2O$ | 100 ml |

stored at $4^oC$

# 6.2 Microinjections

## 6.2.1 Morpholino and mRNA solutions

| Name | Type | Sequence | Reference |
|---|---|---|---|
| zsmoc1sbMO | Splicing | CACATAAATGAAGCGATGAGGGCAT | n/a |
| zsmoc1scMO | Splicing | ATGCATACAGTACCTGAACACACT | n/a |
| zsmoc1tMO | Translation | AAGAGCCAGATTGTGACAGTTCATC | n/a |
| bmp2b MO | Translation | CGCGGACCACGGCGACCATGATC | Lele et al. 2001 |
| bmpr2a MO | Translation | TGTTATTCGGCCTTCAACTGCCATG | Monteiro et al. 2008 |
| bmpr2b MO | Translation | CTGCTGCCGCTTCTGGATGGATTCA | Monteiro et al. 2008 |
| control MO | n/a | CCTCTTACCTCAGTTACAATTTATA | n/a |
| p53 MO | Translation | AAGAGCCAGATTGTGACAGTTCATC | Robu et al. 2007 |

**Table 6.2.1 Details of Morpholino utilised.**

All Morpholinos were supplied as a powder by Gene Tools (USA; **Table 6.1**). Upon receipt MO stock solutions were made up at concentrations of 5mM in ddH2O and stored at 4$^{o}$C. For use each MO was diluted to a working stock (concentrations varied: details in the results). Unless otherwise stated every MO concentration was also co-injected with a p53MO to minimize off target effects caused by p53 dependent morpholino toxicity (Bedell et al. 2011). Before each series of injections the solution was heated to 65°C for 10mins using a PCR machine to resuspend MO. mRNA was synthesised and stored as detailed in 6.3.1. When injecting all mRNA solutions in single use aliquots on ice.

## 6.2.2 Needle preparation

The needles were pulled from the filamentless glass capillaries (Intrafil; dimensions: length-10 cm, diameter: outer - 1 mm; inner - 0.8 mm) using a Micropipette puller (Sutter Instrument, Novato CA, USA) in accordance with the manufacturer's instructions. After being pulled the point of the needle is sealed.

Immediately before use the needle was loaded with 2-3 µls of construct using micro-loader pipette tips (Eppendorf, UK) and then loaded into Picospritzer III micro injector. To break the seal the point of the needle was brought into focus using a Nikon SMZ 1000 stereomicroscope. A micro-manipulator (Intracel, Hertss, UK) was used to finely adjust the position of the needle and the needle was broken with watchmaker's forceps so the needle tip was big enough to penetrate the chorion without breaking, but not so thick that it would lead to unnecessary damage to the embryo, ideally to produce droplets of volume of ~1 nl. The final bolus size was regulated by adjusting the duration of the injection, with pulses ranging from between 10 and 20msec. The size of the bolus was measured on 1000 micron graticule.

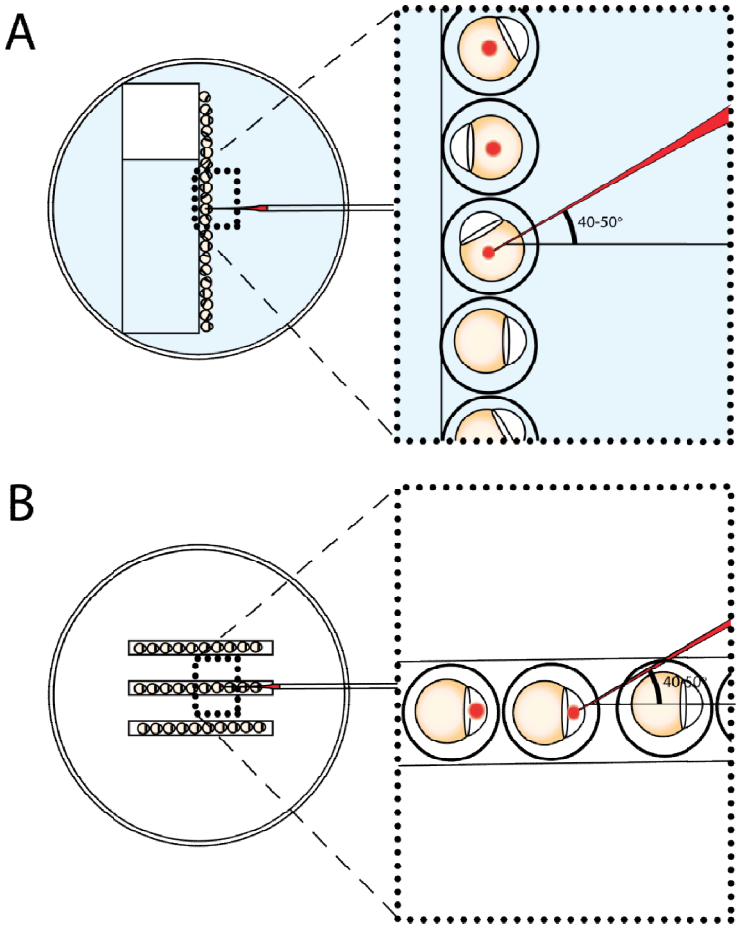



**Fig. 6.2.1 A diagram depicting the two microinjections methods utilised.**

A) Before injecting MO into the embryos they are first arranged next to a histology slides which was wedged into the corner of a 90mm Petri dish. The excess E3 media was removed. This held the embryos in place because of the surface tension of the liquid. Embryos were held in place by the surface tension. The microscope was focused on the topmost embryo and a micro-manipulator was used to drive the needle at an angle of between 40 and 50$^{o}$ through the chorion and into the yolk sac. The dose was then injected. Embryos are injected sequentially by moving the Petri dish vertically. B) In contrast for mRNA injections embryos were arranged in lanes set in agarose set in a 90mm Petri dish from bespoke mould. Using watchmaker's forceps each embryo was aligned so that the cells faced towards the needle while the yolk sac faced away. The needle was then driven through the chorion directly into the cell. The Petri dish was moved horizontally away from the needle allowing the injection of fresh embryos.

## 6.2.3 Microinjections

Microinjections were performed using a Picospritzer III micro injector (Intracel) in combination with a Nikon SMZ 1000 stereomicroscope as detailed in Bill et al. (2009). Morpholino Oligonucleotides were injected in accordance with the methods detailed in **Fig. 6.2.1.** MO were injected into the yolk sac between 1 and 8 cells, whilemRNA was injected directly into the cells of 1 or 2 cell stage embryos. After 6-8 hpf dead embryos were removed and E3 media was replaced.

## 6.2.4 Phenotyping

Embryos were dechorionated **(6.2.3)** then the zebrafish were scored using the Kishimoto scale detailed in **(Fig 3.1.3;** Kishimoto et al. 1997). Embryos were then fixed by overnight incubation in 4% PFA (4°C) before being washed three times in the PBST and stored in 1.5 ml Eppendorf's at 4°C.

# 6.3. Imaging

## 6.3.1 Live zebrafish imaging

### 6.3.1.1 Methyl cellose gel preparation

Unless otherwise stated all zebrafish were embedded in 0.8 % methyl cellulose.

***Master methyl cellose gel (3%)***

| | |
|---|---|
| E3 media | 250 ml |
| Methyl cellose powder | 10 g |

stored at -20$^{o}$C

The E3 media was heated to 80$^{o}$C on a heated magnetic stirrer. The methyl cellose was added whilst it was mixed using a magnetic stirrer. Once the methyl cellose was fully dissolved the heater was switched off and the solution was allowed to cool to room temperature under agitation. The gel was then aliquoted into 50ml falcon tubes (Sterilin) and left rolling overnight. Those not to be used immediately were frozen (and stored for up to 1year).

***Working methyl cellose gel (0.8%)***

| | |
|---|---|
| methyl cellose gel (3%) (RT) | 8.3- 33.3 ml |
| E3 media | 40.2- 15.2 ml |
| penicillin/streptomycin | 0.5 ml |
| Tricaine Solution | 1 ml |

stored at 4$^{o}$C

### 6.3.1.2 Embedding embryos

Embryos were then dechorionated chemically using PronaseE (**6.1.3**). The embryos were then embedded in methyl cellulose in bespoke viewing chambers specifically designed to be loaded into a Prior H101 motorised stage (Prior Scientific Ltd). The details of this process are laid out in **Fig. 6.3.1.**

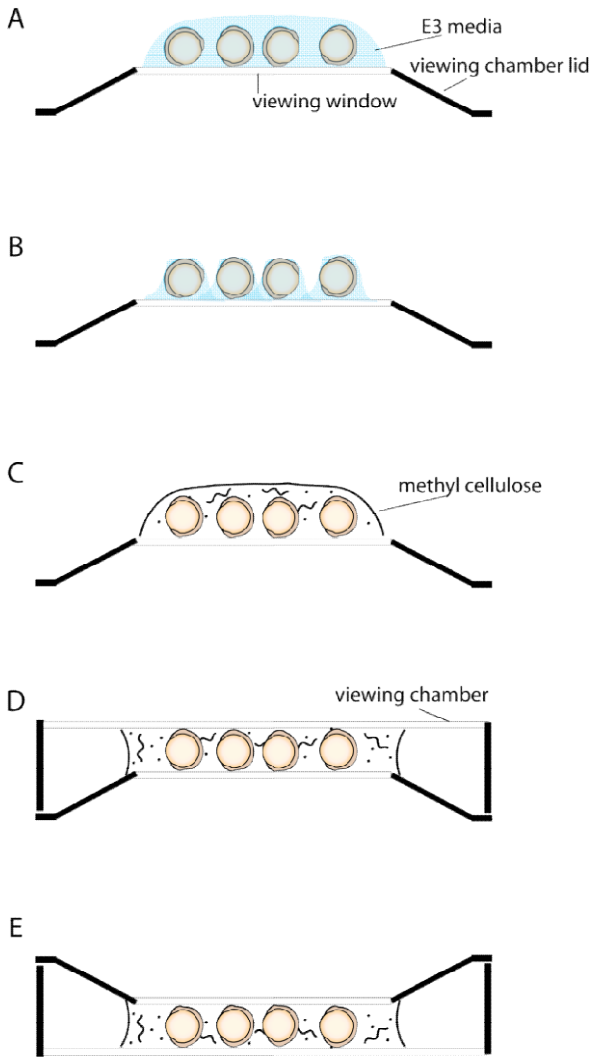



**Fig 6.3.1 A diagrammatic explanation of the embedding of zebrafish embryos.**

**(A)** bud stage embryos were loaded onto an inverted viewing chamber at a density of between 8-15 embryos. **(B)** The excess E3 media was removed. The embryos were still covered with a minimal layer of E3. **(C)** A drop of ~500 µl of methyl cellulose (0.8%) was added and the embryos were held in place by surface tension. Embryos were reorientated and aligned and damaged embryos were removed using a fine tip Pasteur pipette 3.5ml. **(D)** The completed viewing chamber was assembled by placing the viewing chamber on top of the inverted viewing chamber lid. Air gaps in this chamber ensure the zebrafish embryos are not damaged by increases in pressure in the methyl cellulose. The embryos were held in place by surface tension. **(E)** The viewing chamber were then righted and loaded onto the motorised stage.

### 6.3.1.3 live embryo imaging

All live embryo microscopy was conducted on a Nikon AZ100 macroscope (Nikon UK Ltd) in conjunction with a Lumen 200 Pro light source (Prior Scientific Instruments). As the system was used primarily to image reporter lines of zebrafish all images were taken on a Photometrics Coolsnap HQ2 camera (Photometrics Ltd) with Nickon GFP filter cubes. The camera was moved between embryos using an automised Prior H101 motorised stage (Prior Scientific Ltd). The embryos were maintained overnight in an environmental chamber (Solent Scientific) to provide a consistent environment (28.5°C).

### 6.3.1.4 Image capture and analysis

The images were captured using IVision (BioVision Technologies) before being analysed using Fiji (Schindelin et al. 2012). The size of the developing optic field was measured along the axis that bisected at right angles. For details on precisely how these axis were drawn consult **Fig 6.3.2.** The cross-sectional area of eyefield was calculated by manually drawing a region of interest around the optic field and the number of pixels within the region of interest (ROI) was measured. This was converted into $\mu m^2$. To control the poorly orientated embryos or global effects of morpholino treatment the width of the embryo was measured as described in **Fig. 6.3.2** and the total cross-sectional area of the

embryo was measured in the same way as that of the eye. Only in this case the perimeter of the embryo (excluding the yolk cell) was traced as the ROI.

For fluorescent signal quantification a ROI was drawn around the zebrafish optic field (tg(rx3:gfp)) or the area fluorescence (tg(BRE:gfp)). The intensity of signalling was measured by reporting both the mean pixel intensity and modal pixel intensity within the ROI. In the case of the (tg(BRE:gfp) the extent of the GFP gradient was also measured by measuring the width of the widest part of the gradient and then measuring the axis at right angles to this original measurement.

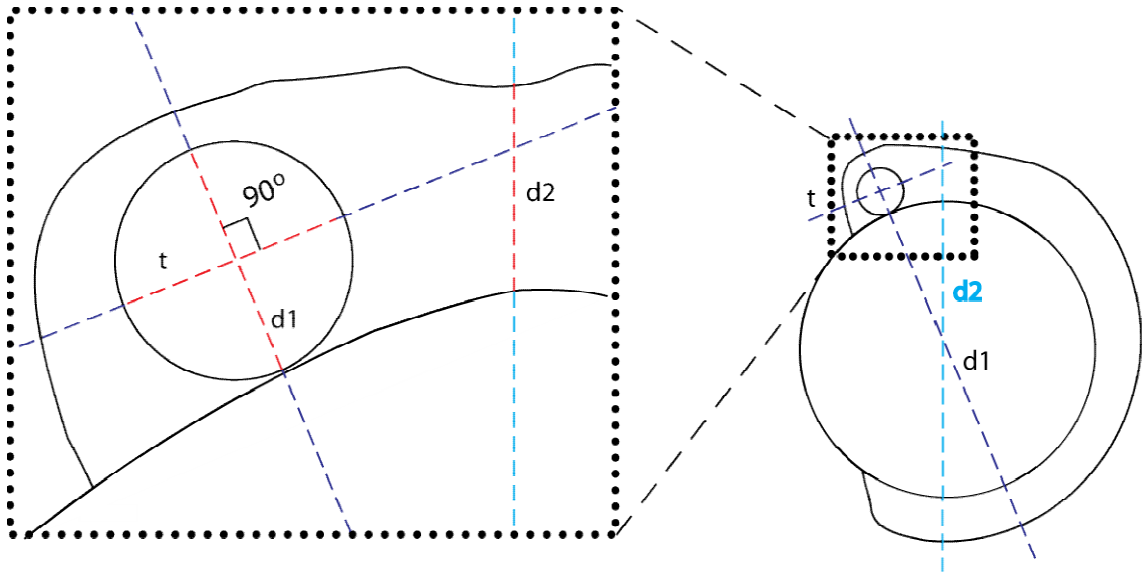



**Fig. 6.3.2 A diagrammatic explanation of the axes used to measure eye field size in live imaging experiments**

The first measurement axis is taken by measuring the width of the eye field along the diameter of the embryo's yolk cell (**d1**). The second axis is taken by bisecting the optic field at a 90° angle from axis (**t**). Axis t is measured at the maximum width of the optic field along this angle. The width of the embryo was estimated by measuring along the diameter of the yolk cell behind the head **(d2)**.

The blue dashed line represents axis while the red dashed line represents the part of the axis that was measured.

## 6.3.2 Gradient quantification with OPT

The embryos were stained by whole mount immuno as detailed in **6.5.5**. Ten embryo of the stained embryos were prepared for Optical Projection Tomography (OPT). The fixed samples were pre-cleaned in PBS, before two further washes in dd$H_2O$. The embryos were then individually embedded in 1% low melting point agarose and dehydrated overnight in Methanol before transferring to BABB (1 part Benzyl Alcohol/2 part Benzyl Benzoate). Each sample was then scanned using the original prototype OPT scanner ( Leica MZFLIII with a Retiga EX camera and a 0.5x objective lens; Sharpe et al. 2002), using the brightfield channel to detect the HrP staining from the whole mount immunos. The raw images were then reconstructed into a 3-D model using the Bioptonics proprietary software suite (Bioptonics). 3-D images were reconstructed from the raw data using VolViewer (Jerome Avondo, JIC, Norwich). The images were then optically sectioned in two planes. To measure the lateral the embryos was bisected 30 μm above the involution zone of the embryo. To monitor the sagittal gradient the embryo was bisected dorsally to ventral. The raw image was exported as a tiff file and analysed using Fiji (Schindelin et al. 2012). To objectively measure the extent of the gradient the images were thresholded at a pixel intensity of 200 using the Fiji colour segmentation tool and a region of interest was drawn around this region manually. The mean pixel intensity and area of this image was recorded. The cross-sectional area of the gradient was initially recorded in Voxels but was

later converted into $\mu m^2$ using calibration data obtained by imaging a 1000 gradual. The data was processed in Excel.

---

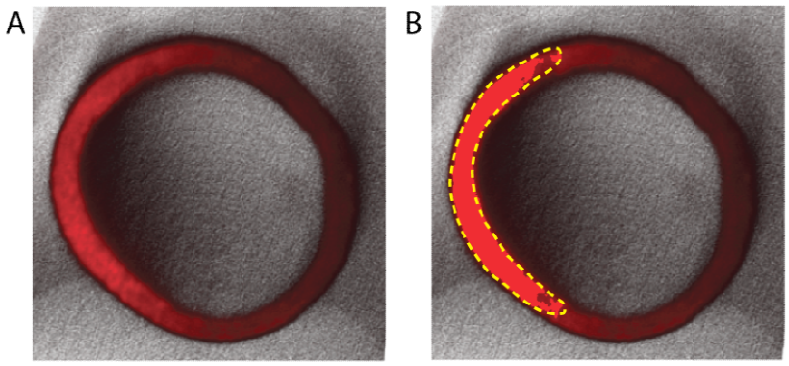



**Fig 6.3.3 An example of gradient thresholding.**

(A) An example of an un-thresholded optical section. (B) the same image thresholded with the Fiji colour thresholding tool. All red pixels with an intensity of over 200 are highlighted. The ROI was drawn around all highlighted pixels.

The yellow dashed line denotes the region of interest (ROI)

---

# 6.4 Nucleic acid techniques

## 6.4.1 General methods and analysis

All DNA solutions were stored in TE buffer at -20°C while all RNA solutions were stored in ultra pure RNase free water (GIBCO). To ensure RNA stability RNA samples were aliquoted for single use and unless they were used immediately samples were stored at -80°C (in the in vitro transcribed RNA was stored for up to 1 month). When is use all RNA samples were stored on ice. To ensure freedom from contamination all solutions were pipetted using filter tipped pipette tips.

To check DNA quality and sample identity samples were always subjected to both spectrophotometric analysis **(6.4.2)** and Agarose gel elctrophosis **(6.4.3)** after any major processing step. Where appropriate sample identity was also confirmed with Sanger sequencing **(6.4.4)**.

# 6.4.2 DNA purification

DNA was commonly purified using commercial spin column based purification. In this case the columns were used in accordance with the manufacturer's instructions. The exception being when RNase free preparations were prepare, in this case phenol-chloroform extraction was used.

**PCR purification**

PCR reaction mixes and some restriction digests were purified using QIAprep PCR purification kit (Qiagen). The DNA was eluted in 50 µL of elution buffer.

**Plasmid extraction**

Plasmids were extracted using spin column kits depending on the desired yield of DNA, detailed in table 6. The extractions were broadly performed according to manufacturer's instructions, the exception being for the QIAprep Maxi kit. The two main differences from the public protocol were that the ultracentrifugation step of the cell lysis buffer was avoided by removing the precipitate by filtration rather than centrifugation and the yields were increased by heating the elution buffer to 65° C before use.

| Kit | Stationary phase bacteria culture | Yield | Elution volume | Manufacturer |
|---|---|---|---|---|
| QIAprep MIni kit | 2ml | ~10 µg | 100 µL | Qiagen |
| Zyppy™ Plasmid Midiprep Kit | 40ml | ~50 µg p | 200 µL | Zymo |
| QIAprep Maxi kit | 200ml | ~250 µg | 500 -1000µL | Qiagen |

**Table 6.4.1 Plasmid purification details**

**Gel extraction**

During cloning it was sometimes necessary to separate digestions products using agarose gel electrophoresis. Once they were resolved the desired band was excised using a scalpel. The DNA fragment was then extracted from the agarose gel using Zymoclean™ Gel DNA Recovery Kit (Zymo) in accordance with the manufacturer's instructions. The DNA was eluted in 7 µl of elution buffer.

**Genomic DNA extraction**

Genomic DNA was extracted from whole zebrafish embryos at 3 dpf using the Nucleon™ HT Genomic DNA Extraction Kit (GE Healthcare) in accordance with the manufacturer's instructions.

**Phenol chloroform extraction**

Phenol chloroform extractions were performed only when the DNA template needed to be completely free of nucleases. The aqueous DNA solution (made up to 200 µl with nuclease free water and 23 µl of 3M ammonium acetate (Ambion)) was initially added to 400ul acidic ultrapure Phenol (Life technologies). The phases were mixed vigorously and separated by centrifugation. 400 µL of analytical grade chloroform was then added to the aqueous phase and again the solutions were mixed vigorously and separated by centrifugation. The aqueous phase was then purified by ethanol precipitation.

To maximise the recovery of the aqueous phase and avoid carrying over phenol or chloroform all separations were conducted using 2ml Manual Phase

LockGel (5 Prime). The entirety of the extraction of a single sample was carried out in a single phase lock tube. After each step the tubes were spun down which allowed the aqueous phase to be separated from the previous extraction with the solid-phase. All extractions were carried out in accordance with the manufacturer's (5 Prime) instructions.

**Ethanol precipitation**

Where appropriate DNA was returned to the aqueous phase via ethanol precipitation. If the manufacturer's guidelines recommended isopropyl precipitation where possible ethanol precipitation was used. The solution was made up of 70% ethanol (using absolute ethanol) and spun at 17000 g for 10 min using a centrifuge called to 4°C. As much of the liquid was removed from the pellet as possible before being replaced with 70% ethanol and spun at 17000 g for 5 min using a centrifuge called to 4°C. If a particularly high purity of DNA was required then the ethanol wash was repeated. The pellets were then air dried and resuspended.

# 6.4.3 RNA purification

**In vivo purification- TRIzol extraction**

Before isolating mRNA in zebrafish embryos were collected (20-30 embryos) immediately and snap frozen using dry ice. The embryos were immediately immersed in 0.5 ml of TRI reagent (Sigma) in a disposable plastic 1.5 ml mortar (Kimble Chase; PN EF24837U) and homogenised with a disposable plastic pestle (Kimble Chase; PN EF24837U). The solution was made to 1 ml by adding a further 0.5ml of TRI reagent (Sigma). The solution was then frozen in dry ice and stored at -20°C for up to 10 hpf. The RNA was then extracted in accordance with manufacturing instructions.

**cDNA synthesis**

Immediately following extraction the mRNA was reverse transcribed. The majority of cDNA was synthesised using the first strand cDNA synthesis kit (Roche; PN: 04379012001). The exception was the cDNA synthesis preceding the RLM race. which was conducted using the reagent provided in the RLM race kit. For specific details consult **6.4.9**.

***First Strand CDNA mix (10-20 µl)***

| | |
|---|---|
| Reaction buffer (x10) | 1-2 µl |
| $MgCl_2$ (25 mM) | 2-4 µl |
| Deoxyribonuclides | 1-2 µl |
| Oligo- p(dt) | 1-2 µl |
| RNase Inhibitor | 0.5-1 µl |
| AMV Reverse Transcriptase | 0.4-0.8 µl |
| total mRNA* | up to 1 µg |
| Nuclease free $H_2O$ | made to volume |

*Before adding the RNA it was incubated at 65°C for 15 min to denature any secondary structure.

The reverse transcription reaction was incubated in a Thermo cycler (DNA engine Tetrad 2; MJ research) using the program detailed in table 6.2.

| Step | Temp | Time |
|---|---|---|
| Annealing | 25°C | 10 min |
| Final extension | 42°C | 60 min |
| Denaturation | 95°C | 5 min |
| Hold | 4°C | up to overnight |

**Table 6.4.2 cDNA synthesis program**

Once synthesised the cDNA was stored at -20°C.

---

**In vitro synthesis of capped and tailed mRNA**

To achieve the maximum levels of in vivo translation the five prime end of the mRNA must be capped with a 7-mythyl guanine and tailed with a string of adenosines. Capped mRNA was synthesised using the SP6 and T7 mMESSAGE mMACHINE kits (Ambion) broadly in accordance with the manufacturer's instructions:

***mMESSAGE mMACHINE reaction (20 µl)***

| | |
|---|---|
| NTP/ Cap Mix (x2) | 10 µl |
| enzyme buffer mix (x10) | 2 µl |
| linearised DNA template | 1 µg |
| nuclease free water | made to 18 µl |
| T7/T3/SP6 enzyme mix | 2 µl |

All templates were made using either CS2+ or the CS2 Gateway vectors. Before the synthesis reaction plasmids were linearised using the protocols detailed in **6.6.2.1**. The antisense control RNA was synthesised by linearising the DNA upstream of the gene and using the reverse promoter (T7 or T3) to initiate transcription. To maximise yield the reaction mix was incubated for two hours using a Thermo cycler (DNA engine Tetrad 2 ; MJ research). Notably the DNase step was omitted, to avoid the introduction of RNases.

To add the Poly(A) tail the Poly(A) Tailing Kit (Ambion) was used.

***Tailing reaction (50 µl)***

| | |
|---|---|
| *mMESSAGE mMACHINE reaction* | 10 µl |
| nuclease free water | 18 µl |
| E-PAP buffer (x5) | 10 µl |
| $MnCl_2$ | 5 µl |
| ATP (10 mM) | 5 µl |
| E-PAP enzyme mix | 2 µl |

The reaction mix was then incubated for 30 min at 37°C on a tetrad Thermo cycler. The RNA purified by the phenol chloroform extraction followed by ethanol precipitation as detailed above **(6.4.2)**, with the exception that basic phenol was used, to exclude the DNA template from the aqueous phase **(6.4.2)**.

## 6.4.4 Spectrophotometric analysis

The concentration and quality of all Nucleic Acid samples was assessed using a NanoDrop 1000 UV-Vis Spectrophotometer (Thermo FisherScientific). The concentration of DNA was determined using absorbance at 260nm. The purity of the sample was assessed by the ratio of the 260nm absorbance to the absorbance at 230 and 280nm. Only samples with a 230/260 ratio greater than 1.5 and a 260/280 ratio of between 1.7 and 2.2 were used.

A low 260/280 (<1.7) value indicates a high level of protein contamination (Thermo Fisher Scientific Inc. 2011). While a high 260/280 (2.2<) indicates extremely high levels on contamination and/or low levels of actual nucleic acid. A low 230/260 ratio implies high levels of contamination, typically carry though from purification (Guanine salts or Phenol; Thermo Fisher Scientific Inc. 2011). Impure samples were either repurified or discarded.

## 6.4.5 Gel electrophoresis

Once a samples Nucleic acid purity was confirmed the quality of the nucleic acid was assessed by gel elctrophosis. All cells were used as analysed by gel electrophoresis and Nucleic acid gels were analysed using gel electrophoresis. This confirmed both that the DNA was good quality (e.g. wasn't degraded) and that the samples had the correct identity (did the fragment migration match the predicted fragment migration).

Unless otherwise stated all gels were run in 2% agarose in 0.5 TBE (w:v). Once the agarose was mixed into the TBE it was weighed and boiled until it had fully dissolved. The solution was weighed again and topped up ddH2O to replace any water loss by evaporation. To stain the DNA ethidium bromide (0.5 μg/ml) was added to the mixture before it was left to set. The samples were mixed with loading dye (6x blue/orange Promega), loaded onto the gel along with either 1kb+ DNA ladder (Invitrogen) or the 100bp DNA ladder (Promega). These acted as reference standard for size migration and positive control. The gel was then run at between 60 and 180V. The gels were visualised with a UV transilluminator.

If the purpose of the gel was a gel extraction the same procedure was followed but the ethidium bromide was substituted with gel and SYBR Safe Blue and the bands were visualised on a blue light transilluminator (Safe Imager 2.0 Blue Light Transilluminator; life technologies) to avoid the DNA damage associated with UV light.

## 6.4.6 Sequencing

Dye terminator sequencing was used to confirm the identity of known nucleic acids or establish the identity of experimentally derived nucleic acids. The sequencing itself was carried out using a 3130/3730 Genetic Analyser (Applied Biosystems) by the Institute of Genetics and Molecular Medicine (IGMM) sequencing service. Sequences reads were either analysed on

sequencher v 4.19.1 (Gene Codes) or ApE. v 2.0 (Paradis et al. 2004). Plasmid maps were also constructed using ApE.

*RNA samples were first reverse transcribed into cDNA and sub cloned into a donor plasmids before sequencing.

## 6.4.7 Ion Torrent sequencing

**Library preparation**

Samples were screened using the workflow detailed in **Fig. 4.2.9**. Embryos (8-20) were injected with either a equal mix of TALEN paired mRNA or control RNA at the one cell stage. Embryos were collected at 3 dpf. The genomic DNA of injected embryos was isolated as described above. The overall strategy for the analyse the TALEN pairs has been discussed at length in **4.2.4.** The region of *zsmoc1* targeted by the TALEN pairs was amplified using the custom primers detailed in **Fig. 4.2.15**. utilising a high Fidelity Phusionflash PCR **6.4.8.** For every TALEN pair tested eight PCR reactions were conducted. Four for the TALEN injected embryos and 4 for the control embryos. As the library is already barcoded because of the custom primers many steps of the standard library prep can be omitted. However the library still needs to be purified. Initially the PCR mix was purified using QAIgen PCR purification columns as described above.

As further purification is required for ion torrent sequencing Agencourt AMPure XP magnetic Beads (NEB) were used to further purify the DNA. The

AMPure beads were left on a roller for 30 min to ensure they were both thoroughly mixed and equaliberated to room temperature. Meanwhile the PCR products were transferred to a 1.5 ml microcentrifuge tube and made up to volume of 160 µl using nuclease free H2O. 280 µl of AMPure bead suspension was then added to each sample and the solutions were mixed by aspiration before being placed on a tube rotator to incubate at room temperature for 10 min.

The tubes were then spun down and placed in a magnetic rack. Once the beads had collected to one side the supernatant was discarded. The tubes were then washed three times in 1.5 ml of 80% ethanol. Each time the supernatant was replaced with 80% ethanol the solution was mixed by rapidly rotating the microcentrifuge tube 180° in the magnetic rack. This placed the magnetic beads on the opposite side of the tube to the magnets in the rack, which led to their migration through the 80% ethanol to the other side of the tube. This was repeated six times for each wash.

After the final wash the supernatant was removed and the tubes were again briefly centrifuged before being placed on the magnetic rack again. This was to allow the removal of residual ethanol. The beads were then air dried for 5 min before being resuspended in 50 µl of TE buffer. To normalise the different amounts of template the samples were quantified with a Bioanalyser (Agilent). If the samples were too concentrated they were diluted at either a 1 in 5 or a 1 in 20 dilution using ultrapure water. Once each sample was accurately quantified they were each diluted to a concentration of 20 pM. All 20 pM samples were pooled to a total volume of ~600 µl (e.g. For 8 PCR products 75 µl

of each sample will be combined to a total volume of 600 µl). The pooled library of samples were then mixed well by aspirating them at least 30 times with a Gilson's pipette (p1000).

**Sequencing run**

The library was then sequenced using a Ion Personal Genome Machine sequencer (Life Technologies). The ion torrent sequencing was performed by the Institute of Genetics and Molecular Medicine (IGMM) sequencing. The sequencing reads were screened for mutations using nextGENe (Soft Genetics) before being exported and tabulated in Excel (Microsoft) using the criteria described in **4.2.4**. Finally to confirm the presence of TALEN induced mutations the number of mutations were compared to the number of mutations found in a control population of embryos injected with antisense TALEN mRNA which have been processed according to the same criteria. A functioning TALEN pair should result in an enrichment of deletions relative to the control because antisense RNA should not induce any mutations.

# 6.4.8 Polymerase chain reaction PCR

### 6.4.8.1 primer design

Before PCR can be performed targets specific primers must be designed. When not using PCR sequences obtained from commercial kits or those published in the literature custom primer sequences were designed. All sequences were designed using Primer3 (Untergasser et al. 2012; Koressaar & Remm 2007). Ideally the primers had a melting temperature (Tm) as close to 55°C as possible though primer Tm ranged between 50-60°C. Each primer was designed to be as close to its partner primer as possible. No primer pairs we utilised with Tm's separated by more than 4°C. To help ensure efficient function where possible the functionality of primer pairs was simulated in silico using amplify 3.1(University of Wisconsin).

The oligos that were not supplied within commercial kits were obtained from Sigma-Aldrich. The specific sequences are detailed in **Sup. Table 7, Sup. Table 8 and Sup. Table 9**.

### 6.4.8.2 DNA amplification

DNA amplification was critical for several experiments, including experiments involving cloning, DNA synthesis and experimental analysis. To

achieve this two types of PCR strategies were used. One utilised Faststart Amplitaq Gold enzyme (Faststart Amplitaq Gold Applied Science) while the other utilised FusionFlashTaq (Finnzymes Fisher scientific). While Q PCR and RACE PCR we utilised purely for analysis.

**Faststart PCR**

When extreme accuracy was not critical (e.g. DNA synthesis and experimental analysis), Faststart Amplitaq Gold enzyme was utilised. The enzyme was obtained premixed a 2x custom mastermix ordered directly from Roche.

***FasttaqMaster mix (x2)***

| | |
|---|---|
| Faststart Amplitaq Gold | 0.12 U/µl |
| Magnesium Chloride (Roche: PN) | proprietary |
| PCR buffer | proprietary |
| dNTPs | 0.4 mM µl |

***FasttaqMaster mix (x1; 12.5-50 µl)***

| | |
|---|---|
| FasttaqMaster mix (x2) | 6.25-25 µl |
| Primer fw | 10 µM |
| Primer rv | 10 µM |
| *DNA template* | |
| Plasmid DNA | 1-10 ng |
| gDNA | 50-250 ng |
| cDNA | 0.2-0.5 ul |
| ddH2O | made to volume |

All PCRs were performed using a Thermo cycler (DNA engine Tetrad 2; MJ research) using the program detailed in table 6.4.3

| Step | Temp | Time | Cycles |
|---|---|---|---|
| Initial Denaturation | 94°C | 5 min | 1 |
| Denaturation | 94°C | 45 s | |
| Annealing | 54-62°C | 40 s | 25-35 |
| Extension | 72°C | 15s+ (15s per kb) | |
| Final extension | 72°C | 10min | 1 |
| Hold | 4°C | up to overnight | 0-1 |

**Table 6.4.3 Fasttaq PCR program**
The dark blue denotes the cycling stage of the PCR program. Every cycle the PCR machine goes through a single denaturation step, followed by an annealing step, followed by the extensions step before returning to the denaturation step until the number of cycles has been completed.

---

**Phusionflash high fidelity PCR**

For experiments where accuracy was critical (principally cloning and the amplification of target DNA for Ion Torrent sequencing) the Phusionflash Taq was used. It was principally obtained premade in the form of Phusionflash 2x Master **Phusionflash** (Finnzymes or Fisher Scientific) mix. So only the primers in the DNA template need to be added before use:

***Phusionflash (25-50 µl)***

| | |
|---|---|
| Phusionflash (x2) | 12.5-25 µl |
| Primer fw | 10 µM |
| Primer rv | 10 µM |
| *DNA template* | |
| Plasmid DNA | 1-5 ng |
| gDNA | 25-50 ng |
| cDNA | 0.2-1 ul |
| ddH2O | made to volume |

After the high fidelity PCR were run using the program shown in Table 6.4.4 on a DNA engine Tetrad 2 (MJ research).

| Step | Temp | Time | Cycles |
|---|---|---|---|
| Initial Denaturation | 98°C | 30 sec | 1 |
| Denaturation | 98°C | 7 s | |
| Annealing | 54-63°C | 5 s | 25-30 |
| Extension | 72°C | 15 s per kb | |
| Final extension | 72°C | 5 min | 1 |
| Hold | 4°C | up to 10 hrs | 0-1 |

**Table 6.4.4 Phusionflash PCR program**
The dark blue denotes the cycling stage of the PCR program. For further details consult **Table 6.4.3**

---

**General**

All PCRs were performed using a Thermo cycler (DNA engine Tetrad 2; MJ research) and the PCR product was purified using the QiaQuick PCR purification kit (Qaigen; 6.4.4).

**6.4.8.3 qPCR**

For experiments that required the accurate assessment of the relative levels of gene expression quantitative PCR (qPCR) was utilised, which relies on the fact that an idealised PCR amplification reaction will result in an exponential doubling of levels of target DNA. Eventually the reactions become saturated, however if the amount of DNA can be measured in realtime it is possible to estimate the absolute levels of the initial template by comparing the number of cycles it takes for the primers to cross a cycle threshold (Ct), which is an arbitrary level of signal above background but before the saturation point. As

long as the same level of intensity is used for each reaction in a run it is inversely correlated to template concentration. While multiple methods are available for convenience comparative Ct method ($2^{-\Delta\Delta CT}$) was used to estimate the relative level of expression. This method assumes an ideal PCR reaction was DNA template doubles every cycle. However no PCR reaction is 100% efficient (Livak & Schmittgen 2001). As shorter amplicons are amplified far more efficiently, qPCR primers were designed to be between 100-200bp long. Additionally to ensure primer pairs were consistent over the concentrations of cDNA we were using they were tested against a concentration curve made from serial dilutions of cDNA. Only PCR primers with efficiencies of greater than 80% were used. Finally the HT7900 machine itself can recognise PCR reactions that do not amplify exponentially and these reactions were excluded.

To quantify the levels of template in realtime SYBR green DNA binding dye (Stragene) was used. SYBR green will form a complex with double-stranded DNA which will fluoresce when stimulated with blue light. SYBR green will fluoresce in the presence of any double-stranded DNA, not just target DNA. To control for this qPCR primers were designed to flank large introns so it was impossible to amplify genomic DNA. Additionally, results were only taken from wells that showed only a single peak in their melt curve which indicates only a single amplicon has been amplified in that well.

As qPCR is an extremely sensitive technique; all PCR reactions were prepared within a PCR room to avoid contamination. Samples were plated out

on 384 well plates (ABgene) in triplicate. To control for variations in template concentration between samples all results were normalised to *elf-alpha*. These genes were chosen as previously published data has shown it to be the most consistently expressed genes over the course of zebrafish development (McCurley & Callard 2008). However, to confirm that treatment wasn't altering the expression of the unrelated housekeeping gene *beta actin* was also measured. If *beta actin* results correlated with *elongation factor alpha (elfa)* then the expression levels of these housekeeping genes was assumed to be independent of morpholino or mRNA treatment.

***Brilliant II Sybr Green Master mix (10 µl)***

| | |
|---|---|
| ***Brilliant II Sybr Green Master mix*** (x2) | 5 µl |
| Primer fw | 10 µM |
| Primer rv | 10 µM |
| *DNA template* | |
| cDNA | 0.1 |
| Nuclease free $H_2O$ | made to volume |

The PCR program detailed in **Table 6.4.5** was run on a ABI Prism HT7900 Sequence Detection System.

| Step | Temp | Time | Cycles |
|---|---|---|---|
| Initial Denaturation | 95°C | 10 min | 1 |
| Denaturation | 95°C | 15 s | 40 |
| Annealing and Extension | 60°C | 1 min | |

**Table 6.4.5 SYBR green PCR program**
The dark blue denotes the cycling stage of the PCR program. For further details consult

Once the data was recorded it was exported into and processed in Excel.

# 6.4.9 RACE PCR

The five prime PCR of *zsmoc1* was isolated using five prime RNA Ligase Mediated Rapid Amplification of cDNA Ends (RLM-RACE) which was conducted with the FirstChoice RLM-RACE Kit (Ambion). The principal improvement RLM-RACE over traditional RACE is the elimination of degraded RNA. To achieve this 10 μg of total RNA was dephosphorylated with a hour-long 37°C incubation with Calf Intestinal alkaline Phosphatase (CIP) which was terminated with a phenol chloroform extraction. This dephosphorylates degraded RNA. The cap of non degraded RNA was then cleaved by incubating ~300 ng of CIP treated RNA with Tobacco Acid Phosphatase (TAP) at 37°C for one hour, leaving one phosphate group on the non degraded mRNA transcripts. This phosphate group enables the ligation of a custom adapter sequence to the five prime end of undegraded mRNA transcripts which provides a known binding site for the five prime primers.

The RNA was reverse transcribed into cDNA using the reagents provided in the FirstChoice kit but with the conditions detailed in section **6.4.4**. The cDNA was then stored overnight for - 20°C.

The cDNA was then amplified using a series of nested PCRs using the Fasttaq mix detailed in **6.4.4.2** but with the PCR program detailed below **(Table 6.4.6)**.

| Step | Temp | Time | Cycles |
|---|---|---|---|
| Initial Denaturation | 94°C | 30 sec | 1 |
| Denaturation | 94°C | 30 s | 35 |
| Annealing | 58°C | 30 s | |
| Extension | 72°C | 30 s | |
| Final extension | 72°C | 7 min | 1 |

**Table 6.4.6 RACE PCR program**
The dark blue denotes the cycling stage of the PCR program. For further details consult **Table 6.4.3.**

---

After each round of PCR the product was run out on agarose gel **(6.4.3)**. After the second round of PCR the bands became visible and the PCR products were TA cloned and sequenced **(6.4.4)**.

All incubations and PCR reactions were performed on our Thermocycler (DNA engine Tetrad 2; MJ research). To avoid contamination all RACE PCR reactions were prepared within a PCR room.

# 6.5 Protein techniques

## 6.5.1 In vitro synthesis

All in vitro protein synthesis of the TALEN constructs were performed using the Retic Lysate IVT™ Kit (Ambion). The mRNA template using the reactions was synthesised and purified using the methods detailed in 6.3.1. The only major modification to the manufacturer's protocol was the omission of radiolabelled methionine in favor of unlabelled D-methionine. As the TALEN constructs are relatively large(~4 kb) the reactions were incubated at 37°C for 2 hours in a water bath.

***mMESSAGE mMACHINE reaction (25 µl)***

| | |
|---|---|
| Nuclease free water | made to 5.75 µl |
| Translation mix (methionine depleted ;20x) | 1.25 µl |
| D-methionine (50 mM ) | 1 µl |
| RNA template (0.5- 2.5 µg) | 1 - 5.75 µl |
| Retic Lysate | 17 µl |

After synthesis the RNA was removed enzymatically with the addition of 2.5 µl of RNase A (Ambion) and the reaction was then incubated for a further 30 min at 37°C. The solutions were then stored for -20°C until use.

# 6.5.2 Western Blotting

### 6.5.2.1 Cell pellet preparation

Before Western blots could commence the embryos must first be lysed to free the protein from constituent cells. Initially 30 embryos were collected at 4 hpf (sphere stage) dechorionated and placed in 1.5 ml microcentrifuge tubes. The embryos were then washed twice in E3 media for 2 min under agitation. Embryos were then resuspended in 100 µl of ringers(-Ca) and the cells were disassociated by aspirating the embryos 15 times using a P200 Gilson's pipette. The solution was incubated at room temperature for at least 5 min before the protease solution was added and the embryos were digested for 15 min at 30°C. After this 40 µl of stopping solution was added to inhibit digestions and the zebrafish cells were pelleted by being spun down at 3500 g for 5 min. The supernatant was aspirated off and the cells were resuspended in 1 ml of cell suspension media. The cells were again pelted by a five-minute centrifugation step at 350 g. The suspension media was removed and the pellet was stored overnight at -20°C.

***Protease solution***

| | |
|---|---|
| Trypsin | 10 ml |
| Versene | 10 ml |

***Ringers (without calcium)***

| | |
|---|---|
| 1M NaCl | 5.8 ml |
| 1M KCl | 145 μl |
| 0.5M HEPES (pH7) | 50 μl |
| $ddH_2O$ | 43.5 ml |

***Ringers (without calcium)***

| | |
|---|---|
| 1M NaCl | 5.8 ml |
| 1M KCl | 145 μl |
| 0.5M HEPES (pH7) | 50 μl |
| $ddH_2O$ | 43.5 ml |

***HEPES***

| | |
|---|---|
| 1M NaCl | 116 ml |
| 1M KCl | 580 μl |
| HEPES (Sigma Aldrick) | 23.83 g |
| $ddH_2O$ | 81.2 ml |

***Stopping solution (x6)***

| | |
|---|---|
| Calf serum | 3 ml |
| 2M $CaCl_2$ | 30 μl |
| PBS (x10) | 1 ml |
| $ddH_2O$ | 6 ml |

Syringed filtered (0.22 μM; Millipore)

***Suspension medium***

DMEM (GIBCO)

### 6.5.2.2 Cell lysis

The cell pellet was resuspended in 50 µl of lysis buffer by repetitively aspirating the solution with a P200 Gilson's pipette. The samples were then incubated on ice for 30 min before being centrifuged at 4°C at 21,000 g for 15 min. The supernatant was then transferred to a clean eppendorf.

***Lysis buffer***

| | |
|---|---|
| RIPA buffer | 925 µl |
| NaF | 10 µl |
| Beta-glycosyl phosphate | 25 µl |
| Phosphate inhibitor mix (x6 Roche) | 40 µl |

***RIPA buffer***

| | |
|---|---|
| ddH2O | 9.5 ml |
| Tris-HCl pH 7.5 (2M) | 250 µl |
| NaCl (5M) | 300 µl |
| 1% NP40 (Sigma Aldrich) | 100 µl |
| 10% SDS | 100 µl |
| Sodium deoxycholate | 50 mg |

### 6.5.2.3 Protein quantification

To normalise the amount of protein loaded between samples the total amount of protein was measured using a Bradford assay. One microlitre of the sample of protein was added to 100 µl of Bradford reagent. The solution was mixed by aspiration and a 2 µl aliquot was taken.. The absorbance was measured at 595 nm using a NanoDrop 1000 UV-Vis Spectrophotometer

(Thermo FisherScientific). The samples were blanked to a control Bradford solution that had only been mixed with lysis buffer. The readings were compared to a standard calibration curve measured by Joe Rainger using known concentrations of BSA.

#### 6.5.2.4 Western blot

All Western blot samples were separated by SDS-PAGE using the NuPage/ XCell *SureLock* system (Life Technologies). Every gel was run out on a precast 10% Bis Tris gel (10 wells ;Life Technologies) at 180 V for ~ 90 min (until the initial loading dye and migrated to the bottom of the gel). Samples were run in reducing conditions in accordance with the manufacturer's (Life Technologies) instructions.

***Sample prep (20 µl)***

| | |
|---|---|
| *Sample in lysis buffer | up to 13 µl |
| NuPAGE LDS Sample Buffer (4X) | 5 µl |
| NuPAGE Reducing Agent (10X) | 2 µl |
| ddH2O | made to volume |

*The western blots on in vitro translation reactions were not normalised. Instead 13 µl of the reaction was loaded into the gel.

Outer chamber: 600ml of 20X NuPAGE MES.
In the chamber: 200ml of 20X NuPAGE MES and 500 µL of antioxidant
500 µl of NuPAGE Antioxidant

Before loading the sample prep was integrated at 75°C for 10 min to denature proteins. To estimate the molecular weight of sample proteins the samples were run with the SeeBlue plus2 pre-stained standard (Life Technologies).

The proteins were then transferred to a Hybond-P PVDF membrane (Hybond-P; Amersham) using NuPAGE® Transfer Buffer (Life Technologies) in conjunction with a XCell II™ Blot Module (Life Technologies). The XCell II unit was run at 30 V for between 45 - 60 min; once the transfer has been completed the membrane is blocked in 5% powdered milk (Marvel; Premier foods) in the PBST for one hour. The blocking solution was replaced and the primary antibody was added (for details consult **Sup. Table 10**) and the membrane was incubated overnight at 4°C. The membrane was then washed in PBST (three times for 5 min) and then incubated with the appropriate HrP conjugated secondary (for details consult **Sup. Table 10**) for one hour. To detect the HrP signal the membrane was first coated with ECL reagents (sourced from either ECL plus or ECL prime kits; Amershame Biosciences). The signal was visualised in two ways. Initially the membrane was exposed to photographic film (Kodak Biomax XAR Film) using a x-ray cassette to provide a physical copy. The length of exposure was determined empirically. Immediately afterwards the signal was visualised using an ImageQuant LAS.

If the membrane was reprobed with the second antibody the membrane was first stripped. This entailed the membrane being washed in PBS before being immersed in stripping buffer.

***Stripping of buffer (1 L)***

| | |
|---|---|
| Glycine | 15 g |
| SDS | 1 g |
| Tween20 | 10 ml |
| $ddH_2O$ | 900 ml |
| Adjust pH to 2.2 with HCL | |
| $ddH_2O$ | made to volume |

The embryo was incubated at room temperature for 10 to 15 min before being washed twice in PBS for 10 min and finally washed twice in the TBST for 5 min. Before the membrane was ready for reprobing it was first blocked in 5% Marvel in the PBST as described above.

## 6.5.3 In vitro digestions of DNA using TALEN pairs

The initial TALEN constructs were obtained using in vitro synthesis (detailed in 6.5.1). The targeted template was obtained by amplifying the target region of *zsmoc1* from p5.1_smoc1using Fasttaq PCR; one of the PCR primers used in the amplification was biotinylated to tag the template DNA. The PCR was performed using the method detailed in **6.5.1**. For information on the specific primers consult **Sup. Table 7 and Sup. Table 9.** The reaction was then assembled as follows:

***In vitro TALEN digestion (20 µl)***

| | |
|---|---|
| In vitro 5'TALEN reaction | 2 µl |
| In vitro 3'TALEN reaction | 2 µl |
| *Restriction buffer (20x) | 2 µl |
| DNA target (5 µg) | 6 - 16 µl |
| dd$H_2O$ | to volume |

*unless otherwise stated the restriction buffer utilised was Roche buffer A.

The reaction was incubated at 28°C eight hours with a Thermo cycler. The solution to be stored at 4°C until use.

## 6.5.4 Southern Blotting

Cleave events were visualised with Southern blotting with the North2South Chemiluminescent Hybridization and Detection Kit (Thermo Scientific). However the methodology required substantial modifications. Initially the DNA was separated using gel electrophoresis **(6.4.3)** on a 3% agarose gel (omitting DNA binding dyes). Simultaneously the positively charged nylon transfer membrane (GE Healthcare) was soaked in SCC buffer (4x). The biotinylated DNA was then transferred to a membrane with a vacuum blot. Once the transfer was complete the DNA was cross-linked to the membrane by baking it at 80°C overnight.

As the DNA template was already tagged the proteins pre-hybridisation and hybridisation typical of southern blots was unnecessary and the steps were

skipped. Therefore after the DNA was cross-linked to the membrane it was initially equilibrated in stringency wash buffer (2X SSC/0.1% SDS) before being washed three times (stringency wash buffer for 15 min each) while being rolled in a hybridisation oven at rt. The membrane was then blocked with a 15 min incubation in blocking buffer (Fisher Scientific). Streptavidin-HRP (Fisher Scientific) was added to the blocking buffer (at a concentration of one part 1 in 300) and incubated with our team for a further 15 min. The membrane was then washed a further three times in wash buffer (x1; Fisher Scientific) for 5 min each time. From then on the signal was detected following the manufacturer's instructions.

The southern blots were imaged in the same method described above for imaging Western blots.

## 6.5.5 Whole mount immunohisto-chemistry

### 6.5.5.1 Embryo collection

Embryos were collected when they reached 60% epiboly. The embryos were then placed in a 1.5 ml Eppendorf and fixed in freshly made 4% paraformaldehyde by incubating overnight while under rotated on a tube rotator. The following morning the embryos are washed in twice PBST (each

wash lasted 5 min). The embryos are then decanted into a Petri dish and manually dechorionated.

The embryos were then dehydrated for long-term storage. This was achieved by transferring them to microcentrifuge tubes and sequentially treated with progressively more concentrated solutions of Methanol:PBST (25% : 75%, 50% : 50% 75% : 25% and 100% : 0%). At each step the tubes were mixed for 5 min on a tube rotator. Once the embryos were suspended in 100% methanol they were stored overnight at -20°C.

### 6.5.5.2 Whole mount immuno methodology

The embryos were first transferred into in situ basket which allowed them to be used in conjunction with a 24 well plate. This simplified solution exchange as baskets were moved between wells. Each well contained 1.5 ml of the appropriate solution. Before use the embryos were first rehydrated. This is achieved by sequential immersion in increasingly concentrated Methanol:PBST solutions (specifically 75% : 25%, 50% : 50% 25% : 75% and 0% : 100%).

Once the embryos were rehydrated they were treated using a modified version of the whole mount immuno found in von der Hardt et al. (2007). Initially the embryos were first blocked for one hour at rt in blocking solution (2%BSA, 2% DMSO, ~0.5% goat serum in PBST), immediately followed by an overnight incubation at 4$^{o}$C in primary antibody (concentration given in **Sup.**

**Table 10**). The embryos were then subjected to 3x1hour washes in blocking solution followed by 3x10 min washes in PBST.

Horseradish peroxidase (HrP) staining of the embryos was achieved by incubating in biotinylated secondary antibody (in PBST) (Vector Laboratories) and washed 3x for 1 hour. The biotinylated secondary antibody was pre-sorbed using 100 - 150 5 day old embryos. These were ground up and added to 10ml PBST along with 100 µl of secondary antibody which was then place on rollers and incubated at rt for 2 - 4 hours.

Embryos were then incubated for 2-3 hours in VECTASTAIN® ABC Reagent followed by 3x1 hour washes in 0.5% Hydrogen peroxidase in PBST. DAB substrate was added and the reaction allowed to continue until staining became apparent, at which point the reaction was stopped by 3x1hour washes in PBS. The embryos were finally stored in 0.4% sodium azide (in PBS) at 4$^{o}$C.

## 6.5.6 Bioinformatic analysis

### 6.5.5.1 signal IP

The translated nucleotide sequence of the five prime of SMOC-1 was analysed using signal IP. The likelihood of a protein containing a signal peptide was assessed using three scores.

### S-score

The S-score is based on the similarity of the amino acid sequence to known signal peptide. This is expressed using a Matthews Correlation Coefficient.

**C-score**

The C score is based on the similarity of the amino acid sequence to known signal peptide cleavage sites. This is expressed using a Matthews Correlation Coefficient.

**Y-score**

The Y score is a combination of both the S and C scores.

All scores were considered insignificant by using a D *cutoff* of 0.50 for the Matthews Correlation Coefficient of the respective scores (Petersen et al. 2011).

**6.5.5.1 PONDR-VLXT**

The intrinsic disorder of the isoforms of the *zsmoc1* was modeled on PONDR-VLXT. This system measures amino acids in nine residue blocks and compares the sequence to published disordered and ordered protein regions. Those that have a Matthews Correlation Coefficient of greater than .5 are predicted to be disordered.

# 6.6 Microbial Methods

## 6.6.1 Basic microbial conditions

All microbial work was conducted on DH5α E. coli cells either sourced commercially (Sigma- Aldrich library competent cells) or from Martin Reijns, with the exception of Gateway destination plasmids. As the plasmids contain the *ccdb* death gene they can only be grown in strains that are resistant to the *ccdb* gene, because of this either One Shot® ccdB Survival™ (Sigma- Aldrich) or DB3 (Reijns) E. coli were used.

All were cultured at 37°C in LB(Luria-Bertani) broth or on LB medium. All bacteria were cultured with the antibiotic that corresponded to the resistance gene(s)( Table 2) at a concentration high enough to maintain selection (**Sup. table 13**). For ease of use cultures were maintained in long-term storage; 450 µl of LB broth was taken from stationary phase cultures and mixed with 150 µl of 90% glycerol. The stocks were stored at -80°C in 2 ml Cryo tubes(screw tops; Greiner). Though all critical plasmids were also maintained as DNA stocks as well.

# 6.6.2 Sub-cloning

The plasmids constructed were synthesised using a range of common cloning techniques ranging from standard restriction and ligation through to Goldengate cloning and finally Gateway Cloning using the methods detailed below. For the information on individual plasmids consult **Sup. Table 11 and Sup. Table 12**.

The majority of the cloning was completed by restriction digests followed by ligations.

### 6.6.2.1 Restriction digests

Restriction digests were performed using enzymes obtained from either Roche or New England bioscience. Enzymes were incubated in the buffers recommended and supplied by their manufacturers. Up to 5 µg of DNA was incubated at 37°C for between one hour and overnight depending on the enzyme. To ensure that the DNA had been completely digested 2 µl of the 20 µl reaction was run out on agarose gels as described above **(6.4.5)**.

### 6.5.2.2 Ligation

Ligations were performed using either Quick ligase (New England Biosciences) or T4 DNA ligase (New England Biosciences) in the associated buffer by incubating the reaction at 16°C for between 2 and six hours.

### 6.5.2.3 Gateway cloning

Gateway cloning is a proprietary cloning system developed by Life Technologies (Invitrogen). It was used to rapidly introduce genes to destination plasmids detailed in supplementary table 2. Initially the genes are amplified using specially designed PCR primers which along with target specific sequence included the attB1 and attB2 sequence on the forward and reverse primers respectively. The DNA fragment was then cloned into the destination vector using the BP clonase:

***BP reaction (10 µl)***

| | |
|---|---|
| PCR product (50-150ng) | 1-7 µl |
| destination vector (150ng) | 1 µl |
| TE buffer pH 8.0 | made to volume |
| BP Clonase II enzyme (Life Technologies) | 2 µl |

The sample was then incubated at room temperature for 60 min before being terminated by adding 1 µl proteins K and incubating the reaction for 10 min at 37°C. Two microlitres of the reaction was then used to transform bacteria (detailed in **6.5.3)** and the plasmid was purified **6.3.1**. Subsequently a donor plasmid was used to clone the gene into the destination plasmid. This was done by taking advantage of the fact that the BP recombination converts the attP sites present in the donor plasmid into an attL site. The attL site can then be recombined with the attR sequences present in the destination plasmid using LR recombinase.

***LR reaction (10 µl)***

| | |
|---|---|
| donor plasmid (<150ng) | 1-7 µl |
| destination vector (150ng) | 1 µl |
| TE buffer pH 8.0 | made to volume |
| LR Clonase II enzyme (Life Technologies) | 2 µl |

The LR reaction was incubated and terminated in using identical conditions to the BP clonase detailed above.

#### 6.5.2.4 TA cloning

TA cloning was conducted using the pGEM-T Easy TA cloning kit (Promega).

#### 6.5.2.5 Goldengate cloning

The theory behind Goldengate cloning is described in **4.1.3.** All the constituent plasmids obtained from the Voytas kit were purified using midi prep kits. Transformations were confirmed by Sanger sequencing.

***First round of Goldengate cloning during TALEN synthesis (20 µl)***

| | |
|---|---|
| RVD vectors(150 ng each) | 8-15 µl |
| pFUS vector (150 ng) | 1-2 µl |
| BsaI (New England Biosciences) | 1 µl |
| Quick ligase (New England Biosciences) | 1 µl |
| T4 DNA ligase buffer (New England Biosciences) | 2 µl |
| dd$H_2O$ | made to volume |

| Step | Temp | Time | Cycles |
|---|---|---|---|
| Restriction digests | 37°C | 5 min | 10 to 30 |
| Ligation | 16°C | 10min | |
| Final digests | 50°C | 5 min | 1 |
| Enzyme denaturation | 80°C | 5 min | 1 |
| Hold | 4°C | 0-1 hours | 0-1 |

**Table 6.5.1 TALEN synthesis program 1**
The dark blue denotes the cycling stage of the program. For further details consult **Table 6.4.2.** The programs run on any Tetrad Thermo cycler.

---

The unligated DNA was digested by incubating the samples for 60 min at 37°C after the addition of 1 µl of both Plasmid-Safe nuclease (Epicentre Biotechnologies) and 10 mM ATP (Epicentre Biotechnologies). Samples were run through a second round of Goldengate cloning.

***Second round of Goldengate cloning during TALEN synthesis (20 µl)***

| | |
|---|---|
| pFUSA vector (150 ng) | 1-2 µl |
| pFUSA vector (150 ng) | 1-2 µl |
| LR-vector (150 ng) | 1-2 µl |
| Destination vector (75 ng) | 1-2 µl |
| Esp3I (Fermentas) | 1 µl |
| Quick ligase (New England Biosciences) | 1 µl |
| T4 DNA ligase buffer (New England Biosciences) | 2 µl |
| $ddH_2O$ | made to volume |

| Step | Temp | Time | Cycles |
|---|---|---|---|
| Restriction digests | 37°C | 10 min | 10 to 30 |
| Ligation | 16°C | 15min | |
| Final digests | 37°C | 10 min | 1 |
| Hold | 4°C | 5 min | 1 |

**Table 6.5.2 TALEN synthesis program 2**
The dark blue denotes the cycling stage of the PCR program. For further details consult **Table 6.4.2.** The programs run on any Tetrad Thermo cycler.

---

After each incubation the desired contract was obtained by using reaction mix to transform competent E. coli (6.5.3) followed by purifying the resultant plasmids using Maxi preps **(6.3.1).**

## 6.6.3 Transformation

All transformations were achieved chemically, either using 50-100 nanograms of plasmids or 1-2 µl of reaction solutions to 50 µl of chemically competent cells **(6.5.1.)**. The cells were then incubated on ice for 30 min before being heat shocked for 30-45 seconds at 42°C. The cells were then incubated again for 2 min on ice before the addition of 450 µl of SOC media. Next the cells were outgrown for 30- 60 min (or 0-30 min for Ampicillin) on an orbital shaker at 37°C. Before being spread out on pre heated (37°C) 90mm LB agar plates made up with the appropriate antibiotic (**Sup. Table 13**). For TALEN synthesis blue white screening was also used. In this case before plating out the transformation 50 µL of 0.1M IPTG (Sigma) and 50 µl X-gal 20 mg/ml (Sigma) was added to the plates. The bacteria were spread and the plates were then incubated overnight and colonies were picked (in the case of blue white selection only white colonies were picked) and grown in LB broth until they had reached the stationary phase and are harvested using commercial kits detailed in **6.4.1**.

# Appendix

## Supplementary Figures

# Supplementary Chapter 1

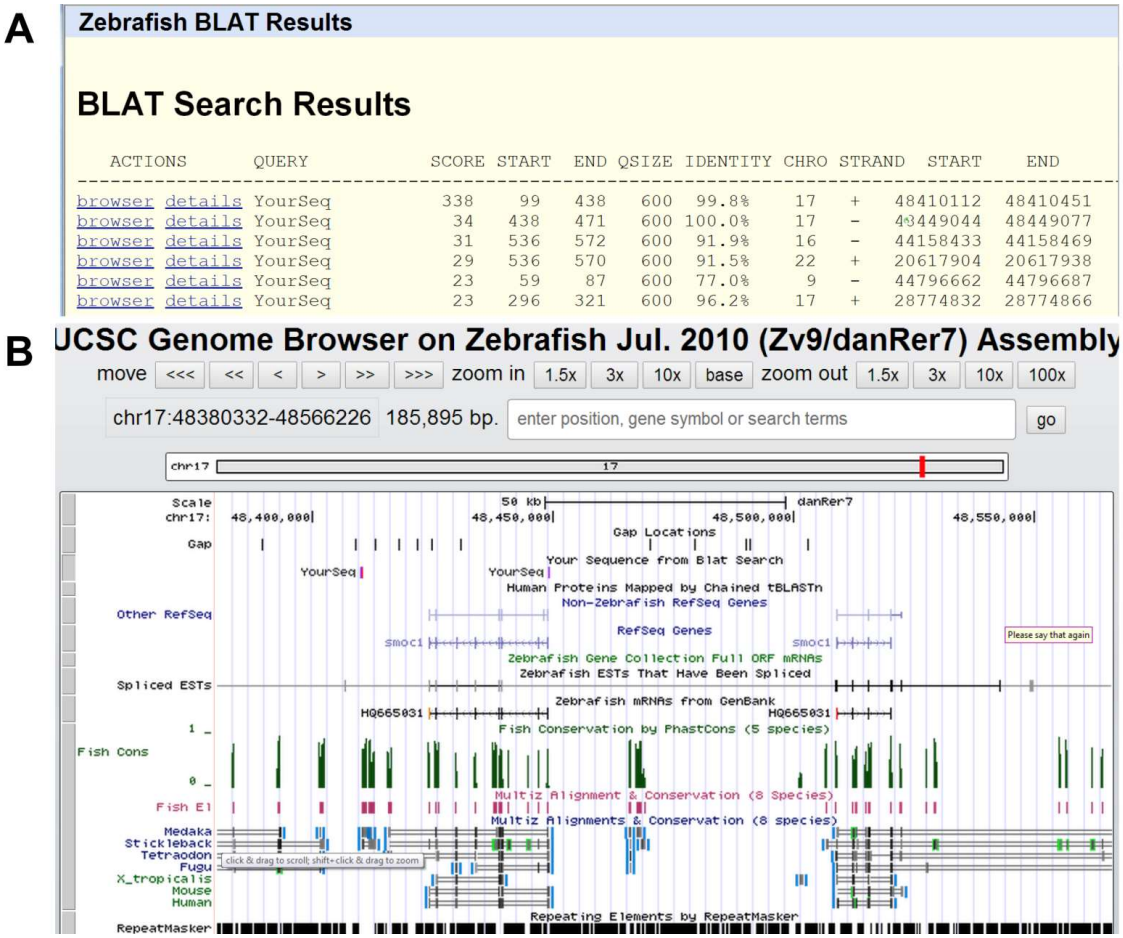
| Defects | Human | Smoc1$^{tm1a/tm1a}$ Mouse |
|---|---|---|
| Ocular | Anophthalmia (92%)<br>Micropthalmia (8%) | coloboma (55.6%) |
| Lower limb | Oligodactyly (83%)<br>Syndactyly (37%) | Oligosyndactyly (91.6%) |
| Upper limb | Syntosis/<br>Syndactyly | None |
| Craniofacial | 37% | 27% |

**Sup table 1. Ocular phenotype summary Rainger et. al. 2011**

# Supplementary Chapter 2

| Gene | Length of intron bp |
|---|---|
| *HSMOC1* | 72,963 |
| *mSmoc1* | 83,192 |
| *xsmoc1* | 58,000 |
| *Mendaka Smoc1* | 74,108 |
| *Drosphilia Pentagone* | 3,387 |

**Sup. table 2. The length base pairs of the first *SMOC1* intron in the human, mouse, *Xenopus* and *Mendaka* genomes along with length of the first exon in *Drosophila Pent* .**




**Sup. Fig 3. UNSC genome BLAT of RACE PCR product sequence.**

**(A)** The BLAT search results of the sequence of 437bp fragment amplified by nested PCR with zsmoc1 targeting primers. The top two blat results correspond to two regions in the smoc1 locus of zebrafish **(B)**. To top it corresponds to a 339 bp
match apparently located the 10kb upstream of the zebrafish loci. The second hit corresponds to the first 33 bp of exon1 of the zebrafish. On zebrafish Jinan version Zv9/danRer7 they are in opposing orientations separated by 30kb. The other BLAT results correspond to sequence originating from the pGEMT plasmid the fragment was cloned into, with the exception of the lowest hit which but the matches are all less than 30 bp long and are spurious

5'UTR

ATTCAACCTGTAGCTCTGGAAAGGACAGTACATCCCACTCAGTGGAAAGAGTGTATCATTGCGACC
ATCACCCTCTTTGTCCGCGTTCTCTCTGGCACTATGAAAACGCACACATGTAAAGGAGAAACCCGTT
GAGGACGAATTTATTTTGGgAACAAACTTTTGGAATTTCCTGCAACGTTGGGA
CTTTTGCCAATTCTTAAGTTAAATCTTGTCCAGTCGTGTTAACACTTCATCGTCAGACAGAATGAACT
GTCACAATCTGGCTCTTTGTGTATTCATGTATTTGATGGTGATGCGCCTGCCAGCAGCCTCAACGCA
GAAATCCACTGGTCCTCG

Exon1

ATGAACTGTCACAATCTGGCTCTTTGTGTATTCATGTATTTGATGGTGATGCGCCTGCCAGCAGCCT
CAACGCAGAAATCCACTGGTCCTCGG

Exon 2

TGGCTTATAGGGGACAGAGAGAGCCCGTGTTCCACGTCCTGCTCCAGAAGTCACGGCAAGCCTGTG
TGTGGTTCAGACGGACGCAGTTATGATACCAACTGTGACCTGGAGAGGGCGAAATGCAGGGACCG
CACGCTCACCCTTGCCCACAGGGGCCGATGCAAAG

Exon 3

GTAAAAACTGGGTGAGGATTGATCAACCTCTTCTTGTGCCCACACTTGTTTCAGAGGTCAAAGAGTT
GACCGGTATGAAAG

Exon 4

AAGCAGGTCAGACGAAATGCCGGACCGAGAGGATTCAGGCTTTGGAACAGGCCAAGAAGCCACA
GGAATCCATTTTCATCCCAGAATGCAACGACGACGGCACATTTGCTCAG

Exon 5

GTCCAGTGTCACACCCTCACAGGATACTGCTGGTGCGTGACGACAGATGGCAAGCCAGTCAGCGGC
TCTTCTGTCCAGAACAAAACGCCAGTGTGTTCAG

Exon 6

GAAACTTCCGTGAATTTATGGGAACTCATGCTGGGACAAGTGGATTGTCAGGGTCAGTA

Exon 7

ACCGATAAGCCACCAGGACCCCCTAGCTCTGGCAAAAAAG

Exon 8

ATGATGGTTCCAAGCCCACACCGACCATGGAGCCGCATGTTGTGCCAGAAGGTGAG

Exon 9

GAGATTACTGCTCCTACATTGTGGATTAAGCAGCTGGTGTACAAAGAGAACAAACAAAACAGTTCG
AATAGTAGGAAGTCAG

Exon 10

AGAAAGTTCCCTCTTGTGACCAGGAGAGGCAGACCGCTCAGGATGAGGCTCGTCAGAACCCGCGG
GAAGCCATCTTCATTCCTGACTGCGGCCTGCAGGGCCTTTACAAGCCAGTTCAGTGCCACCAGTCTA
CAGGTTACTGCTGGTGTGTGCTGGTGGACACCGGGAGACCTATACCTGGAACCTCTGCAAG

Exon 11

GTATAAGAAACCAGAGTGTGATAGTGCAGCTCGCTCTCGAGATACAGAGATGGAAGATCCATTCAG
AGACAGAGATCTGACAG

Exon 12

GGTGTCCAGAGGGAAAGAAAGTGGAATTCATAACAAGCCTTTTGGATGCCCTTACTACTGACATGG
TGCAAGCTATCAACTCGCCAACCCCCTCTGGTGGTGGGAGG

Exon 13

TTTGTTGAGCCTGACCCAAGCCACACACTGGAGGAACGGGTGGTTCACTGGTACTTTGCCCAGCTG
GATAATAACGGCAGTCACGATATCAACAAGAAGGAGATGAAGCCCTTCAAGCGTTATGTGAAGAA
AAAAGCCAAACCTAAGAGGTGTGCCCGCAAGTTCACAGACTACTGCGACCTTAACAAGGACAAGAC
CATCTCTCTTCAAGAGCTGAAAGGATGTCTAGGAGTCAATAAGGAGGG

Exon 14

GTAGCACTACAAGCAGCAGTCAAGGGACAAGGCAAGGGACAAATTTGTTCA

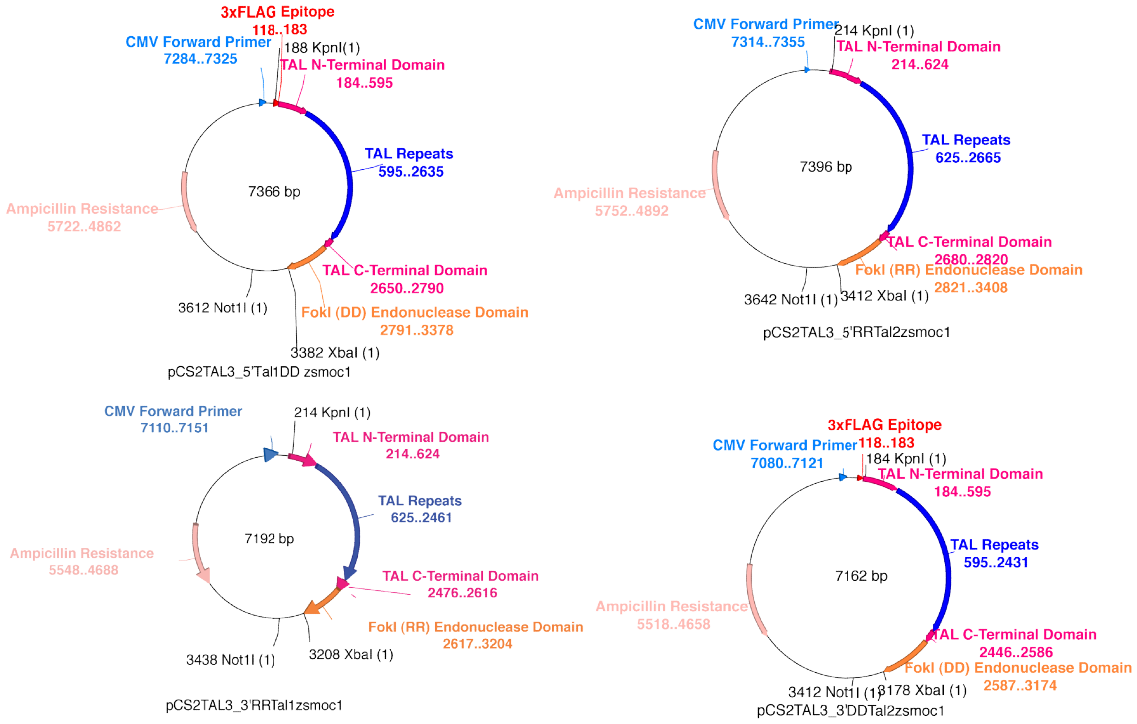
Exon 15

TAGGTCTTCGGGCCTGA

**Sup. Fig. 5. The nucleotide sequence of *zsmoc1* Isoform 1**

# Supplementary Chapter 4


**Sup. Fig. 5. Restriction maps for the pFus plasmids produced in the first round of Golden Gate cloning.**




**Sup. Fig 6. Restriction maps for the completed TALEN constructs**

# Supplementary Methods

## Supplementary Nucleic acid techniques

### Sequencing

**Sup. Table 7.Oligos used for sequencing**

| Name | Sequence | Description |
|---|---|---|
| TalSEQ_fw_1 | CGTGCTGGAGTTTTTCCAGT | Forwards and reverse sequencing primers designed to provide coverage the TALEN protein backbones in pCS2TAL3-DD, pCS2TAL3-DD, pTal3 and pTal4(excluding the RVDs) |
| TalSEQ_fw_2 | ACCGTCCCCTACTTCAGCTC | |
| TalSEQ_fw_3 | CTCAGCTGGAGGGTAAAACG | |
| TalSEQ_fw_4 | GGTGGATCAAGGAAACCAGA | |
| TalSEQ_fw_5 | TTGTTTGTGAGTGGTCATTTCA | |
| TalSEQ_fw_6 | CCAACCGGATAGGGTTCAG | |
| TalSEQ_fw_7 | ATCGCTTCTTGATTCGATGC | |
| TalSEQ_rv_1 | CTGAAGTAGGGGACGGTTTG | |
| TalSEQ_rv_2 | GTTGAATTTCTTGCGATTTCA | |
| TalSEQ_rv_3 | TGCATTTCATCTGCTTGACC | |
| TalSEQ_rv_4 | TTCTCCACCAATCAAAAGCTC | |
| TalSEQ_rv_5 | ATTGCCAAATGTTTGAACGA | |
| TalSEQ_rv_6 | AGCGTGCGTAGATCCACCT | |
| TalSEQ_rv_7 | AACGATGTCTTCGTGTGTCG | |

| | | |
|---|---|---|
| TalSubunitChek_seq_fw | TAGCTGTTTCCTGGCAGCTC | Forwards and reverse sequencing primers designed to provide coverage the RVD's (pNI ,pNN ,pNG and pHD) |
| TalSubunitChek_seq_rv | CACGTTAAGGGATTTTGGTCA | |
| CR8_seq_fw | TTGATGCCTGGCAGTTCCCT | Forwards and reverse sequencing primers designed to flank RVDs in plasmids pCS2TAL3-DD, pCS2TAL3-DD |
| CR8_seq_rv | CGAACCGAACAGGCTTATGT | |
| pTAL_seq_fw | GGCGACGAGGTGGTCGTTGG | Forwards and reverse sequencing primers designed to flank RVDs in plasmids pCS2TAL3-DD, pCS2TAL3-DD |
| pTAL_seq_rv | CATCGCGCAATGCACTGAC | |
| hSMOC1_seq_fw_1 | ACACTGTCCCAGGACTCAGC | Forwards and reverse sequencing primers designed to provide coverage of *hSMOC1*. Primers 1-9 designed by Joe Rainger. |
| hSMOC1_seq_fw_2 | ATCGAGGTCGATGCAAAGAT | |
| hSMOC1_seq_fw_3 | GGCTGTGTTTGTCCCAGAGT | |
| hSMOC1_seq_fw_4 | TTGAGCCAGGGTAATTCAGG | |
| hSMOC1_seq_fw_5 | AGACAGAGTGCCCTGGAAGA | |
| hSMOC1_seq_fw_6 | GCCAGAGCCAAGAGCATAGA | |
| hSMOC1_seq_fw_7 | ACACCCTGGAGGAGCGAGT | |
| hSMOC1_seq_fw_8 | GGTCATCTCACTGCCTGAGC | |
| hSMOC1_seq_fw_9 | CGGACGCCTTGTCTAAAGAG | |
| hSMOC1_seq_rv_0 | AGCATCTTTGCATCGACCTC | |
| hSMOC1_seq_rv_1 | CCTGTGTGAAGGAACCATCC | |
| hSMOC1_seq_rv_2 | AAGAACCACTGATGGGCTTG | |
| hSMOC1_seq_rv_3 | ACACAGGTTGGGTCTCCATC | |
| hSMOC1_seq_rv_4 | GGGGATCACAATGCCCTCT | |
| hSMOC1_seq_rv_5 | GCAACTCCCTGTCCTTGAAA | |

| | | |
|---|---|---|
| hSMOC1_seq_rv_6 | TGGCTGAAGTACCAGTGTGC | |
| hSMOC1_seq_rv_7 | GCCAAGGCTACCACCTTCTT | |
| hSMOC1_seq_rv_8 | GCGGGCACCACTTATGTTAC | |
| hSMOC1_seq_rv_9 | CCAGAAACAAACCACACACG | |
| hSMOC1_seq_rv_10 | GGAGGGTCACGGTCACTTAT | |
| m13_fw | GTAAAACGACGGCCA | Standard sequencing primers used to sequence pGEM-T easy and pDONR221 |
| m13_rv | GTCATAGCTGTTTCCT | |
| zsmoc1fw2 | TTCAGACGGACGCAGTTATG | Forwards and reverse sequencing primers designed to provide coverage of zsmoc1. zsmoc1fw4, zsmoc1fw4, zsmoc1fw2 and zsmoc1rv2 used in the PCR analysis in Fig. 2.2.1. |
| zsmoc1fw3 | CCAGTGTCACACCCTCACAG | |
| zsmoc1fw4 | TGCCAGAAGGTGAGGAGATT | |
| zsmoc1fw5 | AGGGCCTTTACAAGCCAGTT | |
| zsmoc1rv2 | ACTGGCGTTTTGTTCTGGAC | |
| zsmoc1rv3 | TCCTGGTCACAAGAGGGAAC | |
| zsmoc1rv4 | GAGCGAGCTGCACTATCACA | |
| zsmoc1rv5 | GGTCAGGCTCAACAAACCTC | |
| zsmoc1exon13_1_rv | AAGCCACACACTGGAGGAAC | |
| zsmoc1exon13_2_rv | TGTCCTTGTTAAGGTCGCAGT | |
| zsmoc1exon13_3_rv | GACTCCTAGACACCCTTTCAGC | |
| zsmoc1exon13_4_rv | CCCTCCTTATTGACTCCTAGACA | |

# qPCR

| Name | Sequence | Description |
|---|---|---|
| qPCR_efla_fw | CTGGAGGCCAGCTCAAACAT | Forward qPCR primer reference gene amplifies *efla* |
| qPCR_efla_rv | ATCAAGAAGAGTAGTACCGCTAGCATTAC | Reverse qPCR primer reference gene amplifies *efla* |
| qPCR_bactin_fw | TGCGTCTGGATCTAGCTGG | Reverse qPCR primer reference gene amplifies *bactin* |
| qPCR_bactin_rv | TCCCATCTCCTGCTCGAAG | Reverse qPCR primer reference gene amplifies *bactin* |
| qPCR_zsmoc1_all_fw | CAGTTCAGTGCCACCAGTCT | Forward qPCR primer amplifies All known *zsmoc1* isoforms |
| qPCR_zsmoc1_all_rv | CTCTGTCTCTGAATGGATCTTCC | Reverse qPCR primer amplifies All known *zsmoc1* isoforms |
| qPCR_zsmoc1_var1_fw | TGGGAACTCATGCTGGGACA | Forward qPCR primer amplifies only *zsmoc1* isoform A |
| qPCR_zsmoc1_var1_rv | TGGGCTTGGAACCATCATCTTT | Reverse qPCR primer amplifies only *zsmoc1* isoform A |
| qPCR_zsmoc1_var2_fw | CAGGGTCAGTAACCACCGATAAGC | Forward qPCR primer amplifies only *zsmoc1* isoform B |
| qPCR_zsmoc1_var2_rv | TCACCTTCTGGCACAACATGC | Reverse qPCR primer amplifies only *zsmoc1* isoform B |
| qPCR_zsmoc1_var3_fw | TGGGAACTCATGCTGGGACA | Forward qPCR primer amplifies only *zsmoc1* isoform C |
| qPCR_zsmoc1_var3_rv | TGAGGGTGAGAAAGAAACGGAAA | Reverse qPCR primer amplifies only *zsmoc1* isoform C |
| qPCR_otx2_fw | ACCCCTCCGTTGGATACC | Forward qPCR primer amplifies Dorsalisation marker *otx2* |
| qPCR_otx2_rv | ATCCGGGTAGCGTGTTTTC | Reverse qPCR primer amplifies Dorsalisation marker *otx2* |
| qPCR_runx3_fw | GCCCGAAGTCCGAGTACAG | Forward qPCR primer amplifies |

| | | |
|---|---|---|
| | | Dorsalisation marker *runx3* |
| qPCR_runx3_rv | GCTCACCTCCAGCTTCTCC | Reverse qPCR primer amplifies Dorsalisation marker *runx3* |
| qPCR_eve1_fw | TGGATCCTAACGACAAAATGCTA | Forward qPCR primer amplifies Ventralisation marker *eve1* |
| qPCR_eve1_rv | CAGTAGTGAAATACGAGGGCATT | Reverse qPCR primer amplifies Ventralisation marker *eve1* |
| qPCR_gata2_fw | ATCCCGGTACCCTCCTCA | Forward qPCR primer amplifies Ventralisation marker gata2 |
| qPCR_gata2_rv | ACATTCACGCCCCTCTGA | Reverse qPCR primer amplifies Ventralisation marker gata2 |
| qPCR_ID1_fw | AAGAAAGCCAGCAAGATGGA | Forward qPCR primer amplifies Downstream bmp target id1 |
| qPCR_ID1_rv | GCTTCGATGATCTGTTGACG | Reverse qPCR primer amplifies Downstream bmp target id1 |

**Sup. Table 8.Oligos used for qPCR**

# Other Oligos

| Name | Sequence | Description |
|---|---|---|
| fw_RACE_Outer | GCTGATGGCGATGAATGAACACTG | 5' Race primers included in RLM-RACE kit. |
| fw_RACE_Inner | CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG | |
| rv_RACE_zsmoc1_1 | CATAACTGCGTCCGTCTGAA | Three 3' prime nested PCR primers targeted to exon 2 of zsmoc1. |
| rv_RACE_zsmoc1_2 | CAGGCTTGCCGTGACTTCT | |
| rv_RACE_zsmoc1_3 | GAACACGGGCTCTCTCTGTC | |
| rv_RACE_bmp2b_1 | CAGCTCAAGTCGCCTCTTCT | Two 3' prime nested PCR primers targeted to exon 1 of bmp2b as a positive control for RACE. |
| rv_RACE_bmp2b_2 | AGCGCTCAGGTCAAAGACAG | |
| hsmoc1_HindIII_fw | TTTCGTCTCAAGCTTCCTAGTAGACACAGGGTGCCC | PCR primer pair used amplify and Isolate *hSMOC1* by introducing HindIII and BamHI sites flanking the amplicon |
| hsmoc1_BamHI_rv | TTTCGTCTCGGATCCACCGGTACGCGTAGAATCG | |
| zsmoc1fw2_Bioatin | [Btn]TTCAGACGGACGCAGTTATG | Forward biotinylated Primer used to amplify target 4 in conjunction with zsmoc1fw4 |
| zsmoc1rv4_Bioatin | [Btn]GAGCGAGCTGCACTATCACA | Forward biotinylated Primer used to amplify target 4 in conjunction with zsmoc1fw4 |
| zsmoc1SPattb1_fw | GGGGACAAGTTTGTACAAAAAAGCAGGCTGG TTGGGACTTTTGCCAATTCT | PCR primer pair used amplify and Isolate *hSMOC1* by introducing attb1-2 sites HI sites flanking the amplicon for gateway cloning |
| zsmoc1SPattb2_rv | GGGGACCACTTTGTACAAGAAAGCTGGGTA TCAACCCTCCTTATTGACTCCTA | |

**Sup. Table 9. Details of oligos used for cloning, DNA synthesis and RACE PCR.**

# Supplementary Protein Techniques

## Antibodies

| Antibody | Host | Supplier | Product number | Dilution | Protocol |
|---|---|---|---|---|---|
| anti-psmad 1/5/ 8 | Rabbit | Cell signalling | D5B10 | 1 in 200 | Wholemount |
| HrP Anti-rabbit | Goat | Vector laboratories | PK-6200(kit) | 1 in 1000 | Wholemount |
| Anti-FLAG | Mouse | Cell signalling | 2368 | 1 in 500 | Western blot |
| Anti-HA | Rabbit | Sigma | H3663 | 1 in 500 | Western blot |
| Anti-mouse | Rabbit | Dako | P0260 | 1 in 10000 | Western blot |
| Anti-rabbit | Goat | Cell signalling | 7074 | 1 in 10000 | Western blot |

**Sup. Table 10. Details of primary and secondary antibodies used**

# Supplementary Microbial Methods

## Plasmid details

| Name | Details | Selection | Source |
|---|---|---|---|
| pCS2+ | *Xenopus* mRNA expression vector sense expression is controlled of SP6 prompter. The promoter contains antisense T3 and T7. The basic vector also contains a FLAG tag. | Amp | Addgene |
| pCSDest | A CS2 vector containing a gateway cassette. The FLAG tag is absent. | Amp/Chlo/ccdB | Addgene |
| pCSMyc | A CS3 is a variant of the CS2 vector. This vector contains a gateway cassette and a Myc tag. | Amp/Chlo/ccdB | Addgene |
| pGEM-T easy | A component of a TA cloning. The plasmid was obtained pre-linearised with dephosphorylated single base (thiamine) overhangs. | Amp | Promega |
| pDONR221 | Donor plasmids with gateway cassette. Necessary to clone into before the final destination program. | Kan/Chlo/ccdB | life technologies |
| pNI 1 - 10 | RVD plasmid contains the NI rvd site flanked by two BsaI sites. | Tet | Addgene |
| pNN 1 - 10 | RVD plasmid contains the NN rvd site flanked by two BsaI sites. | Tet | Addgene |
| pNG 1 - 10 | RVD plasmid contains the NG rvd site flanked by two BsaI sites. | Tet | Addgene |
| pHD 1 - 10 | RVD plasmid contains the HD rvd site flanked by two BsaI sites. | Tet | Addgene |
| pFusA | First destination plasmid for the first Goldengate reaction. Contains a LacZ gene flanked by two BsaI sites. Which in turn are flanked by two Esp3I. | Spec/LacZ | Addgene |

| | | | |
|---|---|---|---|
| pFusB6-9 | Second destination plasmid for the first Goldengate reaction. Contains a LacZ gene flanked by two BsaI sites. Which in turn are flanked by two Esp3I | Spec/LacZ | Addgene |
| pLR-NI | Last repeat plasmid contains a truncated NI RVD flanked by flanked by two Esp3I. | Spec | Addgene |
| pLR-NN | Last repeat plasmid contains a truncated NI RVD flanked by flanked by two Esp3I. | Spec | Addgene |
| pLR-NG | Last repeat plasmid contains a truncated NI RVD flanked by flanked by two Esp3I. | Spec | Addgene |
| pLR-HD | Last repeat plasmid contains a truncated NI RVD flanked by flanked by two Esp3I. | Spec | Addgene |
| Tal4 | Final destination plasmid for Goldengate cloning of TALEN synthesis. It contains a TALEN backbone . The RVD domains is replaced with a LacZ selection cassette flanked by Esp3I sites. | Amp/LacZ | Addgene |
| pCS2TAL3-DD | Final destination plasmid for Goldengate cloning of TALEN synthesis. It contains TALEN backbone with a modified FokI domain. The RVD domains is replaced with a LacZ selection cassette flanked by Esp3I sites. | Amp/LacZ | Addgene |
| pCS2TAL3-RR | Final destination plasmid for Goldengate cloning of TALEN synthesis. It contains TALEN backbone with a modified FokI domain. The RVD domains is replaced with a LacZ selection cassette flanked by Esp3I sites. | Amp/LacZ | Addgene |

**Sup. Table 11. Plasmids obtained externally for thesis.**

| Name | Details | Selection |
|---|---|---|
| pCS2+_hSMOC1 | *Xenopus* mRNA expression vector(CS2+) with *hSMOC1* cloned in to the Bam1 and HindIII sites. | Amp |
| pCSDest_zsmoc1 | *Xenopus* mRNA expression vector (CS2+) with *zsmoc1* cloned in to the attB1 and attB2 sites. | Amp |
| pCSMyc_zsmoc1 | *Xenopus* mRNA expression vector (CS3+) with *zsmoc1* cloned in to the attL1 and attL2 sites. | Amp |
| pDONR221_zsmoc1 | *E. coli* expression vector (DONR221) with *zsmoc1* cloned in to the gateway cassette. | Kan |
| 5'Tal1zsmoc1_FusA | Destination plasmid for the first ten RVDs of 5'DDTal1zsmoc1 | Spec |
| 5'Tal1zsmoc1_FusB9 | Destination plasmid for the second nine RVDs of 5'DDTal1zsmoc1 | Spec |
| 3'Tal1zsmoc1_FusA | Destination plasmid for the first ten RVDs of 3'RRTal1zsmoc1 | Spec |
| 3'Tal1zsmoc1_FusB7 | Destination plasmid for the second seven RVDs of 3'RRTal1zsmoc1 | Spec |
| 5'Tal2zsmoc1_FusA | Destination plasmid for the first ten RVDs of 5'RRTal2zsmoc1 | Spec |
| 5'Tal2zsmoc1_FusB7 | Destination plasmid for the second nine RVDs of 5'RRTal2zsmoc1 | Spec |
| 3'Tal2zsmoc1_FusA | Destination plasmid for the first ten RVDs of 3'DDTal2zsmoc1 | Spec |
| 3'Tal3zsmoc1_FusB9 | Destination plasmid for the second seven RVDs of 3'DDTal2zsmoc1 | Spec |
| 5'DDTal1zsmoc1 | *Xenopus* mRNA expression vector (CS2+) containing a complete TALEN construct targeting zsmoc1 exon 9. The construct is cloned into CS2TAL3-DD and FLAG tagged. | Amp |
| 3'RRTal1zsmoc1 | *Xenopus* mRNA expression vector (CS2+) containing a complete TALEN construct targeting zsmoc1 exon 9. The construct is cloned into CS2TAL3-RR and HA tagged. | Amp |
| 5'RRTal2zsmoc1 | *Xenopus* mRNA expression vector (CS2+) containing a complete TALEN construct targeting zsmoc1 exon 4. The construct is cloned is CS2TAL3-RR and HA tagged. | Amp |
| 3'DDTal2zsmoc1 | *Xenopus* mRNA expression vector (CS2+) containing a complete TALEN construct targeting zsmoc1 exon 4. The construct is cloned is CS2TAL3-DD and FLAG tagged. | Amp |
| 5'Tal4_Tal1zsmoc1 | *Xenopus* mRNA expression vector (CS2+) containing a complete TALEN construct targeting zsmoc1 exon 9. The construct is cloned is CS2TAL3-RR and FLAG tagged. | Amp |
| 3'Tal4_Tal1zsmoc1 | *Xenopus* mRNA expression vector (CS2+) containing a complete TALEN construct targeting zsmoc1 exon 9. The construct is cloned is TAL4 and FLAG tagged. | Amp |

**Sup. Table 12. Plasmids synthesised over the course of the thesis.**

| Antibiotic | Stock Concentration | Stock Solvent | Working Concentration |
|---|---|---|---|
| Ampicillin | 5 mg/ml | $ddH_2O$ | 100 µg/ml |
| Chloramphenicol | 25mg/ml | Ethanol | 25 µg/ml |
| Gentamycin | 10 mg/ml | $ddH_2O$ | 10 µg/ml |
| Kanamycin | 50 mg/ml | $ddH_2O$ | 50 µg/ml |
| Spectinomycin | 50 mg/ml | $ddH_2O$ | 50 µg/ml |
| Tetracycline | 10 mg/ml | $ddH_2O$ | 10 µg/ml |

**Sup. Table 13. Antibiotic stock solutions**